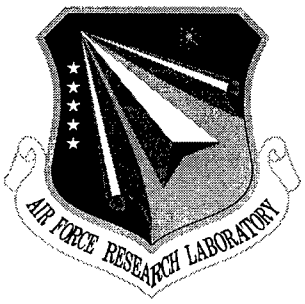
AFRL-IF-RS-TR-1998-56
Final Technical Report
April 1998



VISUAL INFORMATION ENVIRONMENT PROTOTYPE (VIEP)

TASC

Steven L. Rohall

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-1998-56 has been reviewed and is approved for publication.

APPROVED:

PETER A. JEDRYSIK
Project Engineer

FOR THE DIRECTOR:

JAMES W. CUSACK, Chief
Information Systems Division
Information Directorate

If your address has changed or if you wish to be removed from the Air Force Research Laboratory Rome Research Site mailing list, or if the addressee is no longer employed by your organization, please notify AFRL/IFSA, 525 Brooks Rd, Rome, NY 13441-4505. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	Apr 98	Final Apr 94 - Oct 97

4. TITLE AND SUBTITLE

VISUAL INFORMATION ENVIRONMENT PROTOTYPE (VIEP)

5. FUNDING NUMBERS

C - F30602-94-C-0072
PE - 62702F
PR - 5581
TA - 32
WU - 10

6. AUTHOR(S)

Steven L. Rohall

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

TASC
55 Walkers Brook Drive
Reading, MA 01867

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFRL/IFSA
525 Brooks Rd
Rome, NY 13441-4505

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AFRL-IF-RS-TR-1998-56

11. SUPPLEMENTARY NOTES

AFRL Project Engineer: Peter A Jedrysik, IFSA, 315-330-2150

12a. DISTRIBUTION AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report contains information from the work performed under the visual information environment prototype (VIEP) contract to evaluate and integrate technologies to perform collaborative functions involving several different sites, over high speed communication links, and using advanced human-computer interfaces (HCI). A demonstration system was developed to showcase collaborative multimedia technologies in a next-generation C4I scenario, utilizing advanced display and HCI technology.

14. SUBJECT TERMS

Collaborative Multimedia, Human-computer interaction, C4I

15. NUMBER OF PAGES

84

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Standard Form 298 (Rev. 2-89) (EG)
Prescribed by ANSI Std. 239.18
Designed using Perform Pro, WHS/DIOR, Oct 94

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES

1. INTRODUCTION

This report is the last in a series of semi-annual reports documenting the progress of the Visual Information Environment Prototype (VIEP) program. The VIEP program is a 42-month research and development effort funded by Rome Laboratory focused on developing and demonstrating collaborative multimedia technologies for C^3I environments. Under the program, TASC's expertise in collaborative environments, networking, communications, and C^3I is being combined with advances in display and intelligent interface technologies to create a demonstration of next generation C^3I capabilities.

1.1 Motivation

The architecture of the command and control infrastructure in the mid-to-late 1990's will be highly-distributed, supporting cooperative, joint planning among the Military Air Command Center (MACC), Wing Operations Centers (WOCs), and field-mobile Control and Reporting Centers (CRCs). Evolution of this infrastructure is motivated by projections that military conflicts of this decade will be limited-scale, theater-oriented conflicts. Many geopolitical regions exist that have potential for such conflicts.

Meeting the challenge of this defense role requires a C^3I system that enables:

- monitoring of activities and environmental conditions
- archiving and rapid retrieval of tactical imagery
- effective planning, situation assessment and real-time replanning

Such a C^3I system must facilitate the rapid transmission, assimilation and interpretation of information rather than hinder understanding. The VIEP program is examining and using emerging technologies in display, user interaction and collaboration in order to prototype such a system.

Research in the requisite component areas has facilitated this development. Large screen displays have been made larger and less expensive. User tracking and voice recognition systems have been made faster and more accurate. Network bandwidth and image transmission techniques have enabled more rapid transmission of imagery. Collaborative computing has reached a point where it can be used to support network-based conferencing and interaction.

Very little is being done elsewhere to examine the combination and interaction of these various components. Although the individual technologies are able to help users tame the flood of

data, the joining of these emerging technologies will provide the users an innovative environment that will greatly facilitate the control and understanding of data.

The VIEP program is examining and leveraging these emerging technologies, and is combining them with research results from other government-sponsored programs to provide high-performance solutions. For example, under the ImNet program, TASC created a Reliable Adaptive Multicast Protocol (RAMP) allowing for the rapid transmission of information. The advent of reliable multicast has allowed TASC to construct a decentralized collaborative environment with the capacity to add an unprecedented number of users with minimal bandwidth requirements.

Using novel wireless interface technologies, such as laser pointers and wireless gesture recognition capabilities developed at the MIT Media Laboratory, the VIEP program is looking at how new methods of display and user interaction can enhance the utility of collaborative computing. Combining novel interfaces with large screen displays allows users a more natural, real-world feel to the digital information environment. Such a natural feel is deemed critical for the adoption of automated data systems by high-level commanders, where the more familiar computer monitor and mouse interfaces are viewed as too restrictive, cumbersome, and otherwise unacceptable.

1.2 Overview of VIEP Program Tasks

There are three tasks identified under the VIEP statement of work. These tasks are described in the following sections.

1.2.1 Task 1 - Technology Survey

The primary activity of Task 1 is a survey and identification of applicable technologies for integration into the VIEP system. The survey not only served to identify candidate technologies for integration into the first year demonstration, but also identified technologies that showed promise of maturing to a point where they could be integrated into subsequent VIEP demonstrations. Technologies that have been reviewed include: extended, decentralized windowing support for high definition displays (video wall), collaborative human interfaces including novel modalities (e.g. laser pointers, gesture recognition, and speech recognition), visual/multimedia data models, data compression technologies, and multimedia communications services.

1.2.2 Task 2 - Integrate VIEP Testbed

The VIEP program is developing an overall architecture and processing infrastructure, and is integrating technologies identified under Task 1 into the demonstration system. Final function-

ality of the demonstration system will support the focus demonstration developed in Task 3 (VIEP demonstrations). The main output from Task 2 is a design and implementation of the VIEP demonstration system architecture.

1.2.3 Task 3 -VIEP Demonstrations

The overall objectives of the VIEP program demonstrations are to illustrate technologies integrated into the VIEP system and to enable Rome Laboratory to assess progress and appropriate emphasis of the effort. To meet this objective, the VIEP program is developing a demonstration focus scenario to guide system architecture design and choice of functionalities to integrate, as well as to provide yearly demonstrations that illustrate functionality within the context of the focus scenario. The VIEP program is also providing demonstrations of work-in-progress and candidate technologies mid-term of each year of the VIEP effort.

1.3 Report Overview

Chapter 2 describes the focus scenario developed under Task 3 in the statement of work. The VIEP system is being developed to enable collaborative mission planning and real-time replanning. This chapter outlines three demonstration scenarios (routine monitoring, reassigning a mission line, and in-flight replanning/redirection) and uses those scenarios to motivate the development of key technologies areas within VIEP including collaborative computing and novel, wireless user interfaces.

Chapter 3 presents the results of a technology survey completed as part of Task 1 in the statement of work. Technologies reviewed include: pointing systems, spoken language understanding systems, collaborative environments, digital video compression and playback technologies, advanced displays, video conferencing systems, novel wireless pointing systems, and virtual environments. The pros and cons of the various technologies are included and justifications are provided for the particular technologies included as part of VIEP demonstrations.

Chapter 4 details the development and integration of the VIEP testbed as specified under Task 2 of the statement of work. It includes descriptions of 1) the underlying collaborative infrastructure, 2) collaborative conference scheduling and initiation, 3) the operation of the collaborative framework, 4) the individual collaborative tools including image viewer, audio player, MPEG 1 video player and DIVA streaming video player, 5) incorporating third-party applications, 6) wireless interface technologies including speech recognition, gesture recognition, laser pointers and wireless text entry, and 7) wide-area networking capabilities.

Chapter 5 summarizes this report and provides several conclusions concerning the research including the use of collaborative technologies for C^3I, the synergistic combination of technologies within VIEP, and the use of replicated architectures for collaboration.

2. VIEP FOCUS SCENARIO

Although third in the list of Tasks for VIEP, the focus scenario (developed under subtask 3.1) heavily influences the display and interface technologies required (and therefore assessed and evaluated under Task 1). Hence, the focus scenario is presented first to provide context for the discussion of the technology survey and the design and implementation of the VIEP testbed.

In discussions with Rome Laboratory, a main concern for C^3I is bridging the gap between the command operations and emerging digital support technologies, allowing a more natural access to command information. Current display and interface technologies, such as menu-driven graphics displays and pointing devices such as mice, trackballs, and light pens, are substantially more intuitive to use and have greatly facilitated the transfer of meaningful information to the user. Emerging display technologies, including 3D stereo displays and spaceballs further facilitate interaction and information transfer.

A problem with many of the newer input technologies is the requisite direct or indirect tethering of the user to the machine. The prevailing opinion is that high-level commanders will not tolerate such restricted movement or otherwise be so encumbered. Furthermore, working with these devices generally requires a significant amount of training to use them effectively; this type of training is generally viewed as making commanders' jobs harder rather than easier, and as such presents a substantive obstacle for accepting and using these devices in operational systems.

The continuing increase in computational power coupled with novel and improved processing approaches are just now allowing computers to "see," "hear," and "speak to" people directly (albeit in a limited sense) without requiring tethered instrumentation. These "wireless" interface technologies promise to provide the key ingredients that will allow the final bridge between high-level commanders and automated command components. Incorporating these new technologies into a set of applicable tools for C^3I is TASC's primary objective for VIEP. The VIEP vision, showing the use of wireless input technologies, large screen output, and collaborative software is shown in Figure 2-1.

As indicated in the statement of work, VIEP is focused on developing *collaborative* capabilities for C^3I; this is a paradigm shift from the usual hierarchical flow of command from the Military Air Command Center (MACC) through the Wing Operations Centers (WOCs) and the Control and Reporting Centers (CRCs). Nonetheless, there are compelling reasons for providing collaborative tools. For example, late-breaking information garnered at a CRC such as an unsuc-

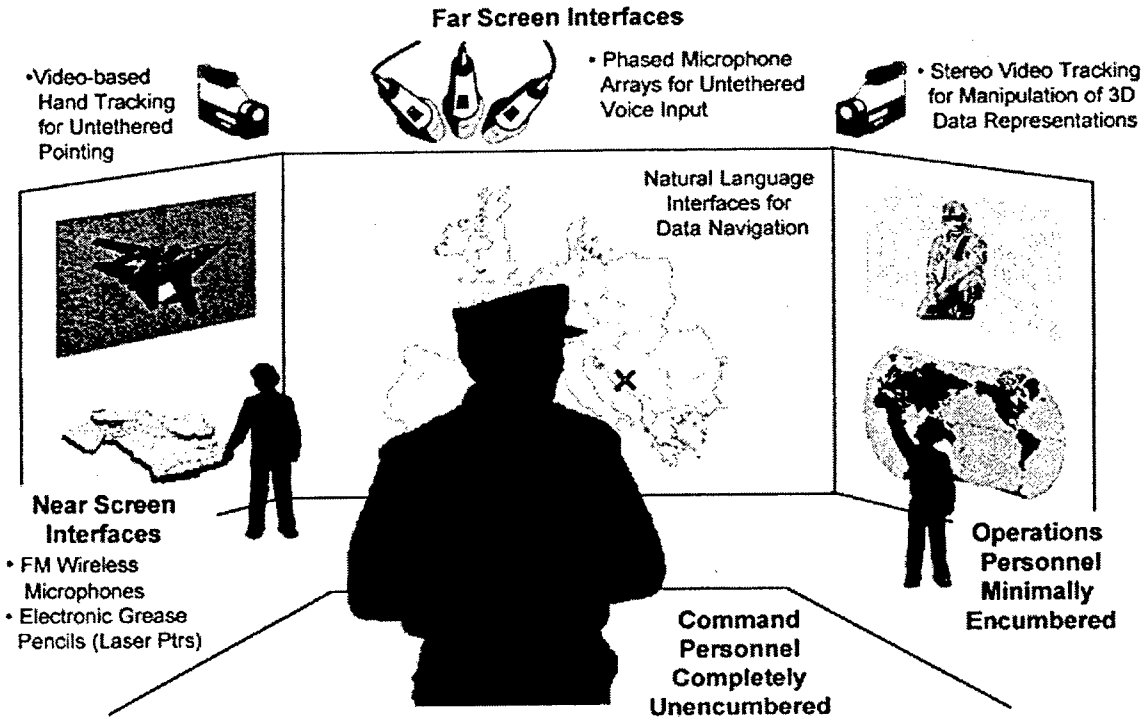


Figure 2-1 VIEP Vision

cessfully completed mission against a high priority target, or a high priority target newly spotted by a mission in progress, should be propagated back to the MACC as soon as possible. Propagating the information is best accomplished in the form of an interactive presentation to the MACC and WOC, both to clarify the information and to discuss options in what is usually a very tightly orchestrated set of mission plans. *If the real-time replanning capabilities needed for interdiction of mobile targets is to become a reality, access to collaborative tools is critical.* Similar to other technological innovations, once collaborative tools are made available we expect that command operations will naturally find increasing uses for these tools.

2.1 Focus Scenario

The intent of the demonstration scenarios is to combine advanced display and intelligent interface technologies with collaborative tools to provide a powerful C^3I capability. In view of the prescribed illustration, an example scenario begins with the display of a large image browser with moving icons representing missions in progress. When one of the missions indicates that its payload has been delivered against a high priority target, gun camera video will be displayed for real-time battle damage assessment. Finding that a target is still operative, another lower priority mission will be redirected in real-time against the higher priority target. The example scenario addresses the year three demonstration objectives for real-time replanning, and allows the incorporation

of wireless interaction technologies including laser pointers, gesture recognition, and spoken language recognition components. Moreover, the scenario incorporates a variety of multimedia data sets including voice, imagery, video, and graphics so that the demonstration will have considerable visual and aural impact as well.

2.2 User Interface Requirements

An example layout for the VIEP user interface is shown in Figure 2-2. (Note that this example layout depicts only the functionality needed to support the focus scenario; the actual "look and feel" of interfaces developed for the VIEP testbed may be somewhat different.) This interface is intended to represent the VIEP user interface that would be present at the MACC and the WOCs. A subset of the windows depicted in the figure would be present at a control and reporting center (CRC) implemented on future generation modular control equipment.

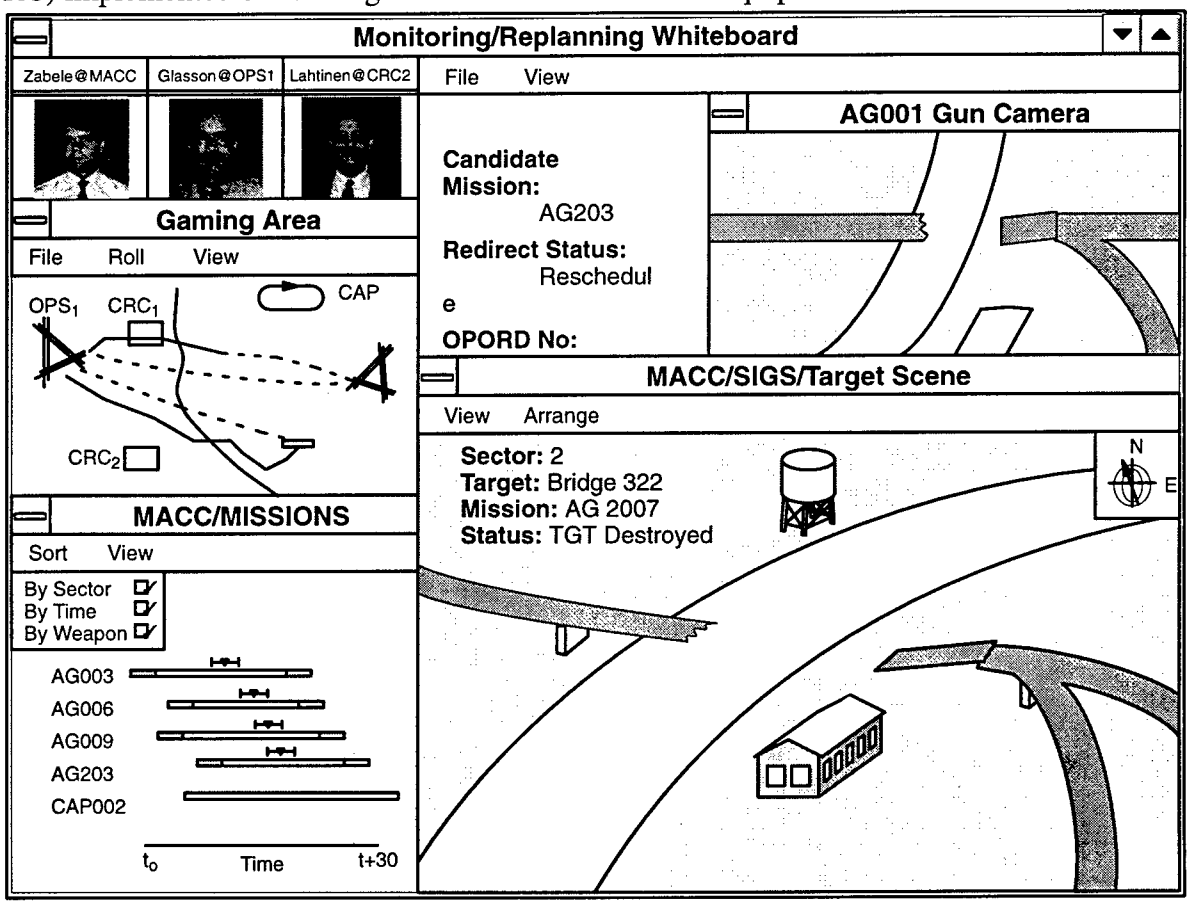


Figure 2-2 Routine Monitoring Scenario

It should be emphasized that the display elements described here are generic visual interface components that have been configured, in this instance, as they would be used in distributed

air operations. These same components would serve equally well in an interface configured for joint operations command, disaster management, or remote medical diagnosis.

The interface is divided into two functional regions. The left region of the interface provides graphical overview information concerning the operations theater, user-selected data such as statistics and planned mission line status, and the conference structure. The right side of the interface is a monitoring/replanning whiteboard that enables the user to display and manipulate artifacts such as mission maps, live imagery, and synthesized renderings.

Items to be placed on the whiteboard can be derived from the left side of the display. For example, a detailed terrain map can be accessed by defining a bounded perimeter on the gaming area window on the left side. Similarly, details of planned and in-progress mission lines could be placed on the whiteboard by designating a line displayed in the MACC/Missions window.

The white board itself could enable display and sharing of multiple information media which could include:

- Free text - such as mission line information cut from the MACC master attack schedule
- Video clips - such as gun camera, radar, or IR sensor imagery
- Local DTED/DFAD maps - to enable evaluation of mission replanning alternatives
- Synthetic and hybrid imagery - to build synthetic images of enemy defense and target areas

All of the media should be accessible via wireless interface. The local DTED/DFAD maps and imagery should also be accessible and manipulable (e.g., viewpoint oriented, zoomed, scrolled) via wireless means.

2.3 Demonstration Scenarios

We now describe a sequence of three example demonstration scenarios with increasing urgency of response, leading to the prescribed scenario requiring real-time replanning. The objectives of the demonstration scenarios are to:

- Demonstrate technology for distributed/collaborative C^3I and wireless interaction
- Demonstrate the potential for dynamic monitoring and replanning in a distrib-

uted environment

- Showcase the utility of visualization for mission review and planning

The scenarios cover the three areas of routine monitoring, reassignment of mission lines, and in-flight replanning. These missions are outlined in the following sections.

2.3.1 Routine Monitoring

The purpose of the routine monitoring scenario is primarily to demonstrate access to the full range of available information in non time-critical situations. Specific features to be demonstrated include:

- Conference scheduling and dynamic entry/exit
- Multicast delivery of information from each of the individual locations such as the MACC and a WOC
- Display of near-real-time video; e.g., gun camera footage
- Display and wireless manipulation of imagery

These features will play a key role in subsequent demonstrations. They also differentiate TASC's distributed/collaborative capabilities from those currently in use at Rome Laboratory.

An example rendering of the MACC's display configuration that could be used for this demonstration is shown in Figure 2-2. On the left side of the layout, the presentation includes:

- Icons representing current conference participants
- The gaming area with traces of ongoing mission lines and theater asset orbits
- A presentation of a segment of the master attack plan mission line

The right side of the display shows video clips from a particular mission platform's "gun camera" (e.g., IR imagery such as the LANTIRN system). To illustrate visualization capability, a synthetic image of the target area is displayed in the lower window. A selectable alternative display for the lower right of the window is a region of the gaming area (designated from the full gaming area map on the left side). Again, the gun camera and target scene images can be accessed and manipulated (scrolled, viewpoint modified) via wireless means. The free text area is unused in this demonstration.

2.3.2 *Reassigning a Mission Line*

In this demonstration, real-time imagery from a mission platform's gun camera indicates that an area of interest has not been successfully destroyed. While the particular target (a bridge) is a high priority target, it is not time critical. The adjustment in response to this damage assessment is to stand down a planned mission and reassign it to the high priority target.

An example display for this demonstration is shown in Figure 2-3. On the left side are the participants, gaming area, and mission marquee. On the right side are:

- The gun camera video indicating that the target has not been destroyed but that the surrounding terrain has been altered by the previous attack
- A synthetic rendering of the target area
- The free text workspace with details of a candidate (substitute) mission line

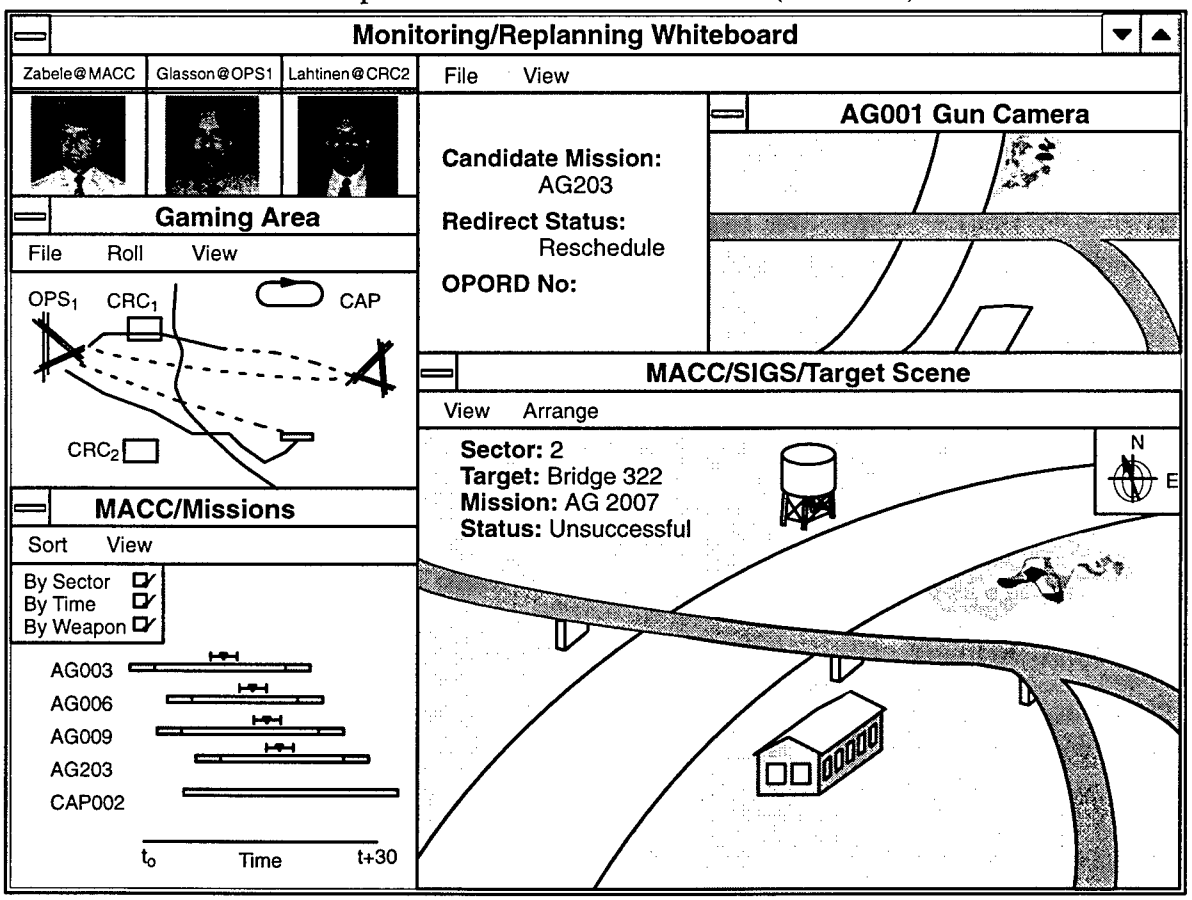


Figure 2-3 Mission Line Reassignment

Manipulation of the target scene could include altering the viewpoint to evaluate the previous mis-

sion's approach azimuth and to explore alternative approaches and initial points (IPs) to improve the likelihood of success of the reassigned mission.

Initially, most of the collaborative interaction in this demonstration would involve a CRC and the WOC evaluating the unsuccessful mission. The focus of collaboration would then shift to the MACC and the WOCs, attempting to identify a lower priority mission line and implement a reassignment. Once the reassignment has been made, the appropriate CRC would be notified.

2.3.3 In-Flight Replanning/Redirection

In this demonstration example, real-time imagery from a reconnaissance platform indicates detection of a time critical target. No mission is currently assigned to this new target; on the order of 10 minutes may be available to destroy the target prior to launch (or take-off). The adjustment in response to this damage assessment is to direct an ongoing mission to attack the time-critical target.

An example display for this demonstration is shown in Figure 2-4. On the left side are the participant icons, gaming area, and mission line marquee. The gaming area display indicates the "racetrack" trace of the TR-1 reconnaissance platform. On the right side are:

- The TR-1 imagery showing a camera aspect of the time critical target
- A synthetic rendering of the target area
- The free text workspace with details of a candidate mission line to be redirected

Manipulation of the target scene for this demonstration could include altering the viewpoint to evaluate favorable approach azimuths. The target area scene on the preferred approach azimuth could potentially be transmitted to the redirected platform as an "on the fly" mission rehearsal aid.

Initially, the collaborative interaction in this demonstration would involve the MACC and a WOC evaluating the reconnaissance information, identifying an appropriate mission and developing an attack strategy. The focus of collaboration would then shift to a CRC to coordinate redirection and to communicate target information with the intended platform. Once the redirected platform has attacked the target, it would then report to the CRC, which would communicate the outcome (possibly with camera clips) to the WOC and MACC.

A demonstration script highlighting the key aspects of the above scenarios has been developed and is reproduced in Appendix A. This script emphasizes the use of VIEP for mission plan-

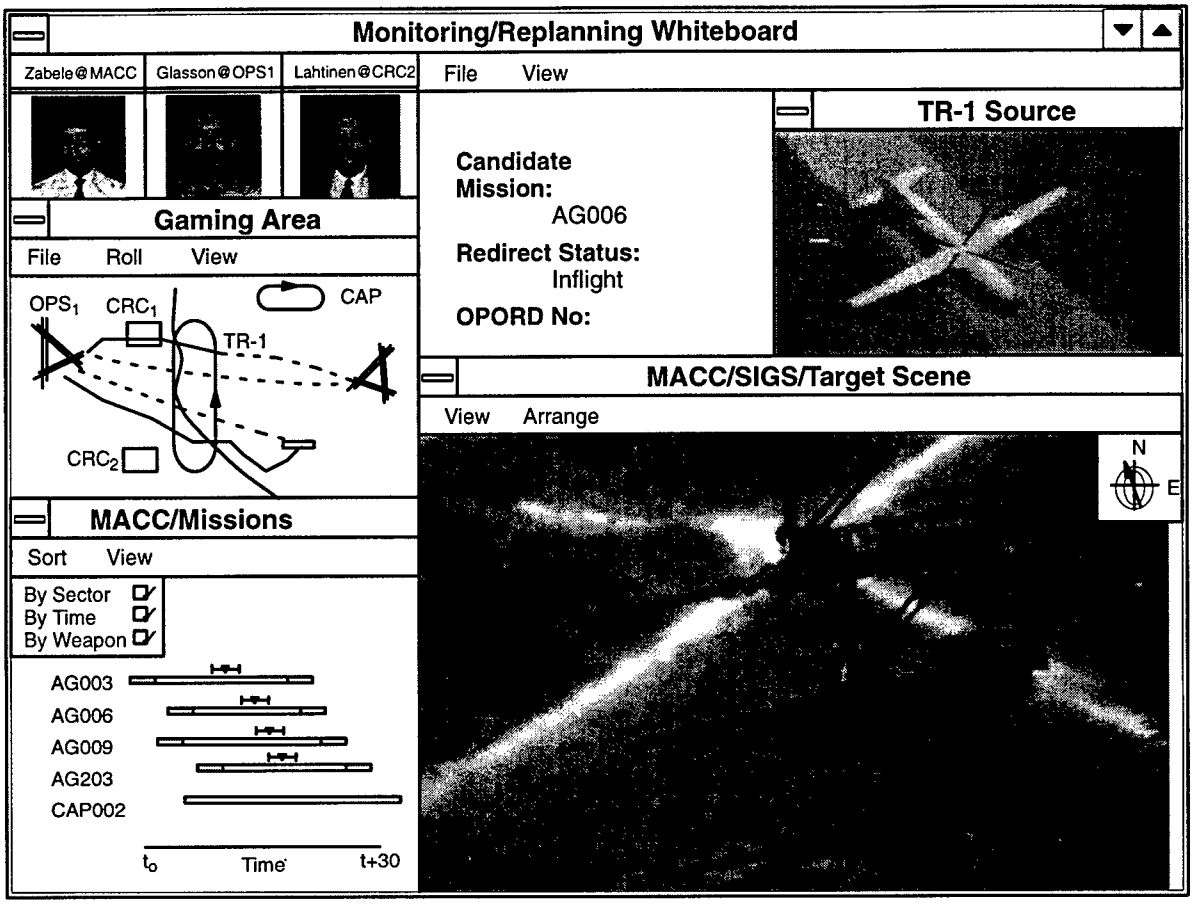


Figure 2-4 Inflight Replanning/Redirection

ning and real-time replanning and has been used for the Year 1, Year 2, Year 3, and Final VIEP demonstrations.

2.3.4 Medical Diagnosis

Although VIEP is being designed and developed with the intent of supporting the C^{3}I scenarios described above, the capabilities of such a system extend beyond the C^{3}I domain. In particular, VIEP could also be used to support a medical diagnosis scenario. For example, field medics communicate with field hospital personnel as well as medical personnel at established hospital facilities. Triage can be done before casualties are moved and, in fact, casualties can be moved to the most appropriate facility for treatment. While en route, X-rays and other diagnostic tests can be performed and shared with surgical staff at the hospital. The doctors have the ability to consult with specialists at other hospitals regarding the incoming casualty before it has even arrived.

The remainder of this document focuses on the C^{3}I domain. It is worth noting, however, that the VIEP testbed has been used for medically-oriented demonstrations.

3. TECHNOLOGY SURVEY

Task 1 of the VIEP program was spent identifying, examining, and evaluating new interface and display technologies for use in VIEP demonstrations. Technologies reviewed include: pointing systems, spoken language understanding systems, collaborative environments, digital video compression and playback technologies, advanced displays, and virtual environments. Subsequent technology surveys were initiated during the second and third years of the program in response to shortcomings discovered in various technology components and in order to add new functionality to the system. In particular, additional wireless pointing systems and video conferencing systems were investigated. The list of technologies evaluated in Task 1 was not intended to be all-encompassing (for example, speaker or face recognition could also play interesting roles in this program), but rather to provide an initial screening of the most promising candidates that would allow the development of an interesting demo to meet the objectives of the program.

3.1 Pointing Systems — Three Dimensional

People naturally want to use their hands when interacting with automated systems, both for specifying (selecting) and for manipulating data representations. Common interface devices, including mice, trackballs, touch pads, touch screens, and graphics tablets, have been adequate for interacting with two-dimensional (2D) datasets and perceptual renderings of three-dimensional objects (so-called 2-1/2D) on 2D displays. Under VIEP, however, there was a possibility of utilizing 3D stereo display capabilities in later stages of the program to provide enhanced access to information such as flight path and terrain information (best viewed in three dimensions). As these conventional interface devices are incapable of tracking movement in three dimensions, novel technologies were investigated. Several interface capabilities are available that allow tracking user movements in three-space including data gloves, spaceballs, "flying" mice, reflective systems, and vision-based ("gesture") recognition systems.

3.1.1 Data Gloves

A data glove (VPL Research's DataGlove and Exos Dextrous Hand Master) fits over a user's hand and relays information about the position of each individual finger. The DataGlove also uses a Polhemus sensor to provide hand position and orientation using an externally generated oscillating electromagnetic field. Nintendo's PowerGlove obtains hand position using sonar devices mounted in the glove; the sonar devices are less expensive but also more restrictive than the Polhemus sensor.

The availability of finger position information from both the VPL and Exos systems is interesting in that it allows the development of a sort of "sign language" for implementing "select" functions as well as other more complex operations. Unfortunately, data glove technology is still too immature to allow such interaction; the number of finger positions that can be successfully and consistently recognized is small, and the system as a whole is not particularly responsive. Data gloves are usually tethered, although untethered forms could be constructed.

3.1.2 Spaceballs

A spaceball (Spatial Systems) is essentially a 3D joystick. A spherically-shaped control is added to the top of the joystick to provide an additional degree of freedom. The device, however, is in essence still a joystick, with the attendant need for a stable base. Although a wireless version is conceivable, manipulating the joystick in mid-air (i.e., without a base platform) would require both hands. Hence a spaceball is ill-suited for use as a pointing system for VIEP.

3.1.3 Flying Mice

A "flying" mouse (SimGraphics Engineering) is a three button mouse with a Polhemus sensor inside. The range of movement that can be supported for the flying mouse is much greater than that for the DataGlove, and positioning accuracy is substantially better. The buttons provide the usual "select" functions associated with standard desk-top mice. Although the expressiveness of button combinations (usually no more than three buttons) is much more limited than that of a data glove's potentially infinite combinations of finger positions, the state of each mouse button is certain and changes are detected and propagated instantly, a tremendous advantage over a data glove. Although flying mice are usually tethered in some form, it is generally easier to create a self-contained unit in comparison with data gloves.

3.1.4 Reflective Systems

A major drawback of each of the above systems is the somewhat bulky and restrictive electronics necessary to provide an untethered environment. An alternative that requires no electronics is offered by reflective IR technologies. Reflective systems (Origin Instruments' Dyna Sight) use a passive IR reflector mounted at key locations (e.g., index finger) coupled with an IR radar system. The receiver is able to resolve the reflector's azimuth and elevation angle, and, when combined the distance information from the radar round trip delay, is able to resolve the reflector's location in three space. This approach is intriguing in that the reflector (~7mm) is both lightweight and relatively unobtrusive. The system is currently unable to support more than one user at a time, howev-

er, which is ultimately problematic for the envisioned environment. Nonetheless, it offers low latency and is a good candidate for inclusion in the VIEP system.

3.1.5 Video-Based Tracking

With continuing increases in computational power and improvements in processing approaches, video-based tracking systems are now becoming possible. Here, either no equipment or simple passive equipment is worn by the user, with an automated system performing hand tracking. By means of a video camera, the computer tracks the positions of the user's hands. Simple chroma key techniques can greatly improve accuracy and performance of such systems although room and clothing color choices may also be affected. To enable such advantages the user might be required to wear a pair of colored gloves; by using different colored gloves, multiple users can be isolated and processed simultaneously allowing the desired multiple user support. Due to resolution restrictions and processing requirements, only the position of the hands can be determined; tracking finger position is still largely unreachable with current technology. Nonetheless, since hand tracking provides the majority of the information for VIEP with minimal inconvenience, it is a good candidate for inclusion in the VIEP system.

Sandy Pentland at MIT's Media Lab has developed a video-based "gesture" recognition system that uses standard SGI video equipment. In fact, low-latency, video-based tracking of hand location and orientation was included as part of the first year demonstration. At the time of evaluation, Sandy's system could track whole body motion at 15 frames per second with 160 by 120 resolution. From discussions with Rome Laboratory, the resolution appeared sufficient for initial demonstrations with two dimensional data sets. In the current system, however, hand position is inferred from simple body articulation models with depth information determined by body position relative to the floor. Consequently, hand positioning in three-space is fundamentally tied to body position, and depth variations of hands relative to the body position is relatively coarse. Use of hand movements to "manipulate" stereo renderings of three dimensional datasets will, at a minimum, require increased depth resolution, particularly if the user wants to stand in one position.

More accurate information can be obtained through a stereo system where depth is determined through triangulation. Use of triangulation provides the added advantage that the whole body need no longer be tracked to obtain depth information. By tracking only hand positions, both processing speed and resolution can be increased. Our initial thoughts on implementation required that the user wear gloves dyed specific colors to allow using a chroma-based detection strategy; however, a simple initialization process could allow the system to determine the color of the users

hands, obviating the need for gloves (if hand color is reasonably separable from the background). At the time of this report, Sandy has completed work on an initial stereo tracking system. This system, though, requires two high-powered SGI Indys to perform the left and right channel video processing. Although work is underway to optimize the system so that a single CPU can be used to process both input streams, doing so would decrease the overall resolution. Having demonstrated the basic technology, we have decided not to integrate the stereo system into VIEP in favor of investigating other wireless pointing techniques.

3.2 Pointing Systems — Near Screen

Despite its compelling advantages, a current limitation of the gesture recognition system is the requisite stand-off distance from the screen needed by the camera in order to "see" the user. Generally the camera is mounted above a large screen display facing the user, and the user interacts with the system from a distance of four feet or more. Consequently there is no mechanism through which the user can approach the screen and "draw" on the screen with his finger as if he had a grease pencil. Grease pencil operations are commonly used for presenting information on large situation displays in command centers, and a digital substitute would provide a familiar capability. Stereo person tracking only exacerbates the situation; the user must be visible to both cameras so the system's field of view is even smaller than in the single-camera situation.

3.2.1 Touch Screens

Large format equivalents of touch screen technologies (Microfield Graphics) offer one solution, where a low power laser scanner is used in conjunction with an array of receivers arranged along the borders of the display screen. By interrupting the beam with a "point" reflector (e.g., a mirrored collar on the tip of a pen), the position can be determined using triangulation. A severe limitation is that the approach supports only one user at a time. When two points are interrupted simultaneously (from two users), the triangulation solution set admits four possible positions, two of which are incorrect. Since supporting multiple simultaneous users is perceived as a necessary capability and this technology is intrinsically not capable of supporting multiple users, this approach is not being pursued.

3.2.2 Light Pens

Another approach involves the use of light pens. Light pens contain light detectors that emit pulses when the display's raster scan sweeps past the position of the pen. By comparing the pulse position from the pen with the timing of the raster sweep (and compensating for processing and

interrupt delays) the location of the pen can be uniquely determined with quite reasonable accuracy. With the addition of a transmitter/receiver pair, a light pen can be made wireless; using orthogonal frequency bands, a wireless light pen approach could easily support multiple simultaneous users. A problem exists, however, in that light pens cannot be used where there is no illumination (i.e., against dark backgrounds) because the sensor has no signal to detect.

3.2.3 Reflective Systems

A third alternative is offered by reflective systems. Although at least two systems would be required to cover a large screen display, the availability of three dimensional information allows a natural automatic "select" function when the reflector comes within a fixed proximity of the screen. Conversely, a deselect (or drop) function could be implemented simply by moving away from the screen. As with the touch-screen approach, the reflective system currently can support only a single user (or at least a single user at a time) and is viewed as a severe limitation. However, discussions with the manufacturer indicate that a multiuser capability is feasible based on the current system, and could be achieved either by using low-power emitters (rather than simple reflectors) or by using frequency-selective reflectors.

3.2.4 Wireless Mice

Another alternative consists of using a wireless mouse that communicates with the CPU via radio frequencies. There are several such devices. Kantek's RingMouse is a clever little device with two buttons worn on the finger. It is tracked in 3-space using triangulation and utilizes infrared signalling to transmit button pushes. Unfortunately, the base unit which does the triangulation is sized for use on a normal desktop computer monitor and not for a large screen application. The GyroPoint is another interesting wireless mouse that uses gyroscopes to sense the motion of a person's hand. The hand's motion is translated into motions of the mouse cursor. Unfortunately, this system will not work if the device is translated in 3-space without any sort of pitch or yaw change. This is problematic since many people hold a pen steady as they write; the GyroPoint system would be unable to see motion in this case.

More recently, devices have been developed to aid people in giving demonstrations. The TrackMan from LOGiTECH is, in essence, a wireless track ball with three mouse buttons. RF is used to communicate the movement of the ball and button state to a base unit. The base unit has built-in support for two wireless devices; each TrackMan can be configured to use a different transmission frequency via a simple switch. For demonstration purposes, this device is very accurate and impressive. The only drawback with this particular device is that the mouse buttons are located

on the same side of the unit as the track ball. This makes it difficult to hold down a button and move the ball with one hand, an operation that is necessary for sketching on a whiteboard or annotating imagery.

3.2.5 Laser Pointers

The best approach appears to be the use of laser pointers coupled with some form of imaging system, either behind the screen (for rear-projection systems) or above the screen (for front-projection systems), to detect and relay screen coordinates. A rear screen system is particularly attractive in that users could actually contact the screen with laser pointers in much the same way that grease pencils are employed today for use on situation boards: both the form factor of the pointers and the lack of tethering ideally suit laser pointers as electronic grease pencil replacements. Furthermore, through spatial separation and appropriate context-switching or even different colored lasers, this approach potentially offers the highly-desirable capability of supporting multiple, simultaneous users.

There is at least one commercially-available system that is based on this idea. The Cyclops system consists of a small camera mounted on an Ovation LCD projection panel combined with laser spot detection and localization circuitry. After an initial set up (where the system is taught the corners of the projected image), the system detects a laser pointer on the projection screen and feeds its location to a Unix host where it is interpreted as a mouse. Since it is possible to use the system in a rear-projection mode, this type of system could address the near-screen pointing capability desired for VIEP. Unfortunately, the attached LCD projector panel does not provide the appropriate quality or resolution for projection of the VIEP system and the system is not available without the costly LCD panel. It is also not clear how multiple users would be handled with this off-the-shelf system.

It appears that the most practical approach is to implement such a laser pointing capability internally. Rome Laboratory personnel have built such a system and it has been incorporated into TASC's VIEP demonstration. The details of this implementation and the changes necessary to accommodate our laboratory environment are described in more detail in Section 4.5.3.

3.3 Speech Recognition/Spoken Language Understanding

A shortfall of the vision-based gesture recognition system is its currently-limited resolution, obviating any real possibility for implementing select functions (e.g., using a sign language-like approach) in the near-term. Consequently, some alternate form of interface is required, at a

minimum, to provide select capabilities. Speech recognition naturally complements the wireless gesture recognition capability in that, if properly configured, users should be able to simply talk to the system to indicate when a graphic object should be manipulated.

A number of speech recognition systems are available, including Dragon System's DragonDictate, Kurzweil Applied Intelligence's Voice system, and IBM's Speech Server Series which includes their Personal Dictation System. The first two systems are speaker-independent (but adapt and improve with use), whereas IBM's system requires roughly an hour's worth of training. All three manufacturers are developing voice dictation systems that provide voice-to-text translation as well as the desired navigation facilities. IBM's software also runs on workstations, which would simplify integration with VIEP components.

The performance of speech recognition systems has been greatly enhanced through the inclusion of natural language processing. By relying on phrase recognition rather than individual word recognition, the recognition accuracy of these systems has increased greatly. The larger benefit of spoken language understanding systems, however, is gleaned through the resulting *interpretation* of the sentence or phrase, in that the meaningless or extraneous words that naturally occur in speech are automatically removed, and physically different phrasings of essentially the same request can be mapped to the same set of actions.

Although our initial requirements are only for limited word recognition to implement select functions, a spoken language understanding approach offers still greater potential for enhanced demonstrations. For example, rather than simply responding to "select" and "release" commands, the system could be made to respond to more abstract queries, such as "show me all missions within twenty minutes of the target." Consequently we have constrained our evaluation to systems providing spoken language understanding support as it provides an evolutionary path for improved demonstrations during the course of the program or beyond.

There are many competing spoken-language systems including Apple's PlainTalk, IBM's Continuous Speech Series, BBN's HARK system, SRI's Decipher system, Speech System's PE400, and Verbex Voice System's Listen. Several other "development" systems are also available, including work being done by Victor Zue at MIT LCS. IBM's CSS, BBN's HARK, SRI's Decipher and Victor Zue's system operate on workstations and are therefore more amenable for integration into a C^3I demonstration. The other systems are constrained to Intel-based PCs or Macintosh systems, which is still problematic for operation in a multitasking environment. Also these other systems are, in general, not as sophisticated and do not have the recognition performance.

The two alternatives that we specifically evaluated are BBN's HARK system and Victor Zue's spoken language system at MIT LCS. A comparison of the two systems reveals that each provides both speech recognition and language understanding capabilities. As such, each supports our initial requirements for simple voice commands (e.g., "select", "release") needing only speech recognition functions, as well as longer term requirements for supporting more generalized voice commands (e.g., "overlay the current weather imagery," "show me all missions within twenty minutes of this target"). Although Victor Zue's spoken language system appears to have a slight advantage for handling simpler word and phrase recognition tasks, the HARK system has the more substantive advantages of being a supported, commercial product with development tools. We therefore concur with Rome Laboratory's selection of the HARK system for advanced UI development.

3.4 Advanced Displays

Although the amount of data available to users has increased dramatically, this increase has not been followed by a corresponding increase in the number or effectiveness of tools to aid users in understanding the data. A simple method for helping users manage the increased data volume is to provide larger display areas, both in terms of resolution and in terms of sheer physical size. Coupled with three-dimensional graphics and stereo presentations systems, large screen displays provide dramatically increased effectiveness for presenting and communicating complex information sets. As such, we intend to use large screen displays as part of the VIEP program. While initial demonstrations of the VIEP system utilized two-dimensional display, the use of stereo in conjunction with the large displays also remained an option. Therefore, many of the options described below can work both in a non-stereo and in a stereo environment.

We have been examining various large screen display options for TASC's node in the collaborative demonstration. Rome Laboratory's ADII Laboratory is already equipped with a large display capability that is suitable for VIEP demonstrations.

3.4.1 Virtual Reality Helmet

A so-called virtual reality (VR) helmet provides both stereo presentation capabilities and the effect of a large screen display. By using a separate, independent display for each eye, excellent stereo separation is achieved, and by locating moderate-sized screens in close proximity to the eyes the apparent size of the display area is increased. A Polhemus sensor attached to the helmet is often used to provide spatial location information that is used to control the images presented to the user in near-real-time. By adjusting the images to correspond to the current viewing direction, the effect

of a considerably larger screen is created. Although a helmet provides convincing presentations, it is both cumbersome and currently requires tethering. Although the tethering could be eliminated, the bulkiness and weight of the system remains a problem. As our investigation is focusing on a natural, untethered environment, helmet-based solutions are unacceptable.

3.4.2 SGI Presenter

SGI is currently offering a color, high-resolution LCD projection system called the SGI Presenter. Output from an SGI monitor is routed through the Presenter and projected onto a wall or similar reflective flat surface using an overhead projector as a lighting source. Using (possibly corrective) mirrors, the Presenter could probably be adapted for use in a rear-screen projection system (a requirement for VIEP), however the Presenter is ill-suited for use in stereo viewing. Although the refresh rate, quality, and resolution of the Presenter are acceptable for normal display operations, the refresh rate is not fast enough to provide acceptable stereo images using SGI's field interleaving capability.

3.4.3 Projection Systems

There are a number of higher bandwidth, albeit less portable, systems based on projection systems that can support stereo viewing. Projection systems include both front projection and rear projection types. A front projection system is located behind the users, with light from the projectors striking a reflective surface in front of the users where the image is formed. A rear projection system is located in front of a user behind a semi-opaque screen where the image is formed, and is similar to projection televisions found in homes. A major focus of this project is novel user interactions, most of which require that the user stand in front of the screen. This therefore eliminates the use of a front projection screen as the user's presence creates shadows in the viewing area.

There are two common methods for creating color 3D stereo with large screen displays. The first of these two methods makes use of polarization techniques, where the right and left images are sent through twin projectors, each projector emitting light polarized orthogonal to the other. (Alternately, systems have been constructed using a single projector with a lens switching polarization in synchrony with the left and right field components; however, there is a substantial loss of illumination with this approach which is deemed unacceptable.) Each user wears a pair of glasses where the right lens has the same polarization as the right projector and the left lens has the same polarization as the left projector. In theory, as the right projector is polarized orthogonally with respect the left projector, the right eye receives no light from the left projector. In practice, however, since polarization is imperfect, some light from each projector reaches the opposite eye.

Through careful alignment, this "cross-talk" can be reduced sufficiently so as to not interfere with the fusing of the two images, and generally provides the best separated and brightest image suitable for common viewing. Use of polarized stereo systems with rear-projection systems requires an expensive non-depolarizing screen that, coupled with the requirement for two projectors, renders this approach quite costly.

A second common method involves rapidly displaying alternating left and right stereo fields in conjunction with active glasses (SGI CrystalEyes) worn over the eyes that alternately block out each eye in synchronization with the video display. If the switching rate is sufficiently high, the display appears unbroken. Within the CrystalEyes system, liquid crystal screens alternately switch between transparent and opaque states: while the left lens is transparent the right lens is clear and vice versa. The receiver on the glasses and a transmitter at the projector synchronize the images. Since the pixel decay rate may be greater than the switching rate, cross-talk may occur here as well. Generally cross-talk is much smaller than polarizing systems and is easier to correct. Reduction of cross-talk can be achieved by slowing down the exchange rate between images; however, this must be weighed against distracting flickering effects that occur when the frame rate is too low. For normal use with SGI's 120Hz field interleaved display system and comparably-equipped projection systems (with the exception of the SGI Presenter), cross-talk is not a problem. While active glasses are still somewhat obtrusive, this approach requires only one projector and a standard rear projection screen. As the cost of the glasses is relatively small in comparison to projectors and non-depolarizing screens, this approach is considerably less expensive for small audiences and is preferable for implementing a large screen at TASC.

3.5 Video Compression

In order to make available all media types relevant for C^3I, digital video support is also being provided under VIEP. Video provides planners with new types of information. Reconnaissance and archived video can be used in threat assessment before a mission as well as for evaluating mission success. Gun camera and reconnaissance video can be used during or after a mission for damage assessment and performance evaluations. Simulations and weather data loops can be retrieved and evaluated. Video can also be used for monitoring civilian broadcasts. In a collaborative environment, video allows different parties visual feedback and video representations of other parties. By combining stereo cameras, 3D displays and network transport, one is able to create 3D video conferencing capabilities.

Use of digital video requires compression both to minimize storage requirements and to

provide more efficient use of network resources. A substantial number of video compression options is available, including Indeo, QuickTime, H.261, CELL-B, MPEG 1, and MPEG 2. For C^{3}I applications, the selected approach should provide the highest possible quality (feasible for near-term demonstrations) so that analysis functions (such as battle damage assessment) as well as stereo perception (for 3D teleconferencing) are not impeded by image degradation. Indeo and QuickTime primarily provide simple, software-based video capabilities for personal computers, and do not provide the performance (quality vs. compression ratio) that other approaches provide. H.261 (and other video conferencing standards) are geared more towards video conferencing where backgrounds are assumed not to change significantly between frames. This restriction is likely to be problematic for gun camera types of video sequences. CELL-B and other vendor-specific hardware solutions are not (and will likely not become) standards, and are available only on limited hardware sets that do not include SGI equipment.

Motion JPEG, as well as JPEG "variants" including MPEG 1 and MPEG 2 are industry standards that provide exceptionally high quality, and either have or will have hardware acceleration for encoding in the near future. As MPEG 2 implementations, as well as MPEG 2 hardware acceleration, were not available, we selected MPEG 1 as the common solution for current and future use.

We currently have access to a PC-based MPEG 1 hardware encoder that we have used to generate MPEG 1 compressed video sequences. We have also obtained a copy of Boston University's MPEG 1 software decoder, and have successfully demonstrated collaborative playback of the compressed MPEG 1 sequences on an SGI Indy. The BU implementation is a derivative of the much acclaimed Berkeley MPEG 1 implementation, with the addition of real-time playback support. Real-time playback provides frame skipping, such that frames are presented concurrently with audio playback to achieve audio and video synchronization. This allows slower machines that are incapable of full frame rate decoding to maintain audio-video synchronization, and more importantly for collaborative applications, to maintain video synchronization across multiple platforms. In the final year of the program, we incorporated streaming MPEG video capabilities developed under the Rome Laboratory-sponsored DIVA project.

3.6 Computer Supported Collaborative Work (CSCW) Environments

The Computer Supported Collaborative Work (CSCW) component provides the basic infrastructure for integrating the various interface technologies and providing the collaborative C^{3}I capabilities. Ideally, the CSCW system should support a large number of users who can be widely

distributed from each other. The system should also provide good performance in spite of large network latencies and the use of large sets of multimedia data. Finally, the system should be extensible and provide an Application Programming Interface (API) that allows both integration of new technologies and development of new application types.

Several commercial offerings are available that provide CSCW capabilities in the form of desktop conferencing systems. These systems generally provide network-based voice and video transport as well as some form of shared whiteboard facility that can be used for exchanging images, text, and graphics. Offerings in this area include SGI's InPerson, IBM's Person-to-Person, Sun's ShowMe, Hewlett-Packard's MPower, Intel's ProShare, InSoft's Communiqué!, InVision's Desktop Conferencing System, AT&T's Vistium, PictureTel's PCS, Microsoft's NetMeeting and many others.

As these systems are all relatively new offerings, there is a widespread lack of standardization, i.e. these systems are generally not interoperable. H.320 is an International Telecommunications Union - Telecommunications Standard Section (ITU-T) group of standards for teleconferencing that includes the H.261 video codec standard as well as several others. This standard is not well suited for software implementations, nor does it support document conferencing (shared whiteboards). As such, several of the larger companies including Intel and IBM are currently promoting their own individual standards. Although several of the other vendors are beginning to adopt these standards (AT&T has agreed to support Intel's ProShare standard by transcoding), standardization issues will likely not be resolved for several years.

Irrespective of the interoperability issues, the greatest obstacle for adopting one of the commercial systems for use under VIEP is the general lack of support for SGI equipment. As the common infrastructure element for VIEP is SGI (and in particular, the Rome Laboratory display system currently requires the use of SGI equipment), selection of an SGI-compatible solution is imperative. Of the numerous offerings in this area, only the SGI InPerson system currently fills this need. One other vendor (InSoft) now has an SGI port of their product, but it was not available in time to meet Year 1 program demonstration requirements.

There is also a general lack of support for third party integrators due to the absence of APIs. In particular, SGI's InPerson does not provide an API, nor have plans for one been disclosed. InSoft has, however, recently made available their OpenDVE API under version 4.0 of Communiqué!. As such, InSoft currently offers the most likely long-term solution for a commercial system. Choice of InSoft as a provider would also play well against InSoft's installed based within the DoD

(for example, all SSCN nodes including the one at Rome Laboratory, are using Communiqué!) and offers an excellent opportunity for technology transfer of VIEP-developed components.

In order to accommodate near-term demonstration objectives, we elected to use the CSCW environment developed by TASC under the DARPA-sponsored ImNet program. As this system was developed on SGI equipment, and we control all of the source code, the system is ultimately extensible. Moreover, the system was designed from inception to provide the needed features of scalability, flexibility, and performance that is currently unmatched by any of the commercial offerings. The advantages of our CSCW infrastructure include:

Scalability - A primary goal of our CSCW design effort was the creation of an architecture that supports a large number of users, and that supports efficient exchange of large multimedia data volumes, such as image and video data sets. CSCW applications are traditionally implemented using unicast protocols which function well for two or three users; however, performance degrades quickly because bandwidth demand increases as the square of the number of users. When the inter-user messages consist of multimedia data, the potential exists for complete application failure. As a consequence, collaborative applications built upon unicast protocols are quickly mired with even moderate numbers of users, particularly for multimedia applications. We have therefore emphasized the use of multicast protocols whenever possible and appropriate, as the required bandwidth will increase at most linearly with the number of users. With multicast, a message need only be transmitted once, yet it can be received by all conference participants. To achieve this end, our CSCW communications model centers on the reliable multicast protocol developed by TASC for the DARPA-sponsored Image Networking (ImNet) project.

Performance - A second goal in formulating the CSCW architecture was optimizing the performance of interactive access and manipulation of large multimedia data volumes. The typical approach for image retrieval, for example, used both in commercial document management systems and in shared whiteboard applications requires that the entire image be transferred across the network before it can be displayed or manipulated. This approach is infeasible for large images. As an example, a high-resolution scanning of a standard size chest x-ray can yield over 100 Mbytes of data. Even at ideal ethernet speeds, transferring the entire x-ray would take well over 90 seconds. Further, a typical workstation would not have sufficient virtual memory to view the entire image.

Our objective was to support high-speed interactive access to such large multimedia data volumes by providing only the data necessary to support each users' immediate requirements. For example, interactive access to the digitized x-ray can be achieved by presenting the user with a

reduced-resolution overview image that allows the user to scroll and zoom to the point of interest. Scrolling and zooming cause the retrieval of additional image data across the network, but only the amount of data needed to fill the user's display window is retrieved. Only the data that is actually needed is ever transmitted across the network.

Flexibility - A third goal was designing a CSCW architecture that supports the breadth of human interactions typically encountered at meetings and conferences. Traditional CSCW architectures constructed using shared windowing systems, such as Hewlett Packard's SharedX and Farallon's Timbuktu, use a What-You-See-Is-What-I-See (WYSIWIS), interaction model where all users see the same data and all manipulations have global effect. Typical human interactions, however, do not fit this model. For example:

- Within a single conference there can be multiple side sessions that an individual moves among
- Within a session there are often short, spontaneous side discussions between members from the same organization in addition to any central discussions being held by the group as a whole
- An individual will often privately examine material within conference proceedings or briefings other than what is currently being presented in order to clarify a previous point or to preview upcoming material
- An individual will occasionally contact a colleague not involved with the conference to discuss the consequences of any newly presented information or ask for additional information before making a presentation.

Our objective was to support a full range of interaction modes, including the ability to switch between public and private control of views and the ability to set up subconferences where a subset of the conference participants can share and manipulate information without affecting the group as a whole.

Eliminating mandatory global synchronization of all views has the added advantage that users can independently control the presentation and layout of screen area, thereby allowing different users to have different views open or to have views arranged differently on the display. As applications become more complex and competition for screen area increases, independent control of screen layout by individual users is extremely desirable.

The design goal of supporting a wide range of interaction modes coupled with the design goal of supporting a large number of users offers an additional means for optimizing image conferencing performance. By eliminating the mandatory global synchronization of all views and al-

lowing users to select and manage views independently, the particular data sources that each user is actively engaging are explicitly identified. Combining this information with the explicit group setup properties of multicast, the transport of individual sets of image data can be restricted to only those users currently needing the data. The combined approach offers substantial gains in conferencing performance, even over systems such as Bellcore's Rendezvous that support multiple simultaneous user interactions but remain tied to TCP/IP protocols.

Diversity - A final goal was designing an architecture that readily supports a full range of multimedia data types, including images, graphics, text, audio, and video, as well as an architecture that is extensible so that new data types, interaction modes, or manipulation tools can be integrated quickly and easily as conceived or become available. The ability to incorporate specially-designed tools not only improves system performance, but also facilitates system development.

3.7 Audio and Video Conferencing

The VIEP system assumes that collaborators have audio and perhaps video communications with each other. In particular, the TASC-developed CSCW infrastructure does not implement any sort of floor control among collaborators. Instead, it relies on communications among the collaborators to coordinate their actions. This is akin to the human-to-human protocols that naturally develop as part of an audio teleconference, for example.

All of the commercial packages described in the above discussion on CSCW environments include audio and video conferencing capabilities. Although these packages provide nice, integrated interfaces, only one of them (SGI's InPerson) utilizes multicast for efficient network usage during multi-party conferences, only one other (InSoft's Communiqué!) provides a programmer's API for controlling the audio and video, and all of them are expensive, ranging in cost from several hundred to several thousand dollars per seat.

Given the use of our own CSCW infrastructure, the need for a nice, integrated audio and video package was diminished—separate audio and video tools would appear integrated within our infrastructure. As a result, we investigated several free packages developed as part of the MBONE project: Lawrence Berkeley Laboratory's VAT for audio and Xerox PARC's NV, Inria's IVS, and LBL's VIC for video. All of these packages support the necessary functionality, utilize multicast (enabling the demo to scale), and include access to the source code (should customization of some sort be necessary). We chose to incorporate VAT and VIC into VIEP because they were both written by the same organization. As a result, they have similar user interfaces. In addition, VIC can be configured to perform voice-activated switching in conjunction with VAT. When VIC is con-

figured to follow the active speaker, the video decoding and display requirements are greatly reduced, especially in a multi-party conference.

4. VIEP TESTBED INTEGRATION

The goal of Task 2 is the design of the VIEP demonstration system. Using component technologies identified in Task 1, a testbed was developed to demonstrate the utility of collaborative multimedia for command and control. This section describes the development of the prototype system. It includes detailed descriptions of 1) the underlying collaborative infrastructure, 2) conference scheduling and initiation, 3) the operation of the collaborative framework, 4) the individual collaborative tools, 5) incorporating third-party applications, 6) wireless interface technologies that have been incorporated into the system, and 7) wide-area networking capabilities.

4.1 Enhanced Collaborative Infrastructure

VIEP testbed development began with a reworking of TASC's CSCW infrastructure per the original decision to use our own CSCW infrastructure for VIEP development. At the time the initial technology survey was performed, it was the only solution that met our criteria for scalability, performance, flexibility, and diversity. In addition, it was the only alternative to run on an SGI workstation and have an API available (in this case, access to all of our source code). InSoft had developed an API for their Communiqué! product, but, at the time, Communiqué! only ran on Sun workstations.

The original CSCW architecture (Figure 4-1), designed and built as part of the ImNet program, consisted of monolithic, tightly-integrated applications. Applications communicated with each other using the TASC-developed RAMP reliable multicast protocol, so they were very network efficient. However, the monolithic structure made it difficult to incorporate new services, e.g., a new MPEG viewing tool. In addition, it was difficult to support platforms with different capabilities, e.g., not all collaborators may want or need video support if they have a low-bandwidth connection to the rest of the conference.

The new CSCW architecture (Figure 4-2), resolves the problem by separating the collaborative tools such as image and MPEG viewers from the core communications component. Messaging between collaborators occurs via RAMP messaging between each collaborator's core communications component. The core communications component routes messages as needed to individual *installable modules*, e.g., image or MPEG viewers. This new structure simplifies integration of new tools, reduces core application size, and results in faster launch and execution. The IPC protocol between the core communications component and the installable modules allows individual components to be added, updated, or replaced without affecting the rest of the system. It also allows for simple "stubs" to be provided for sites missing specific components. Finally, our

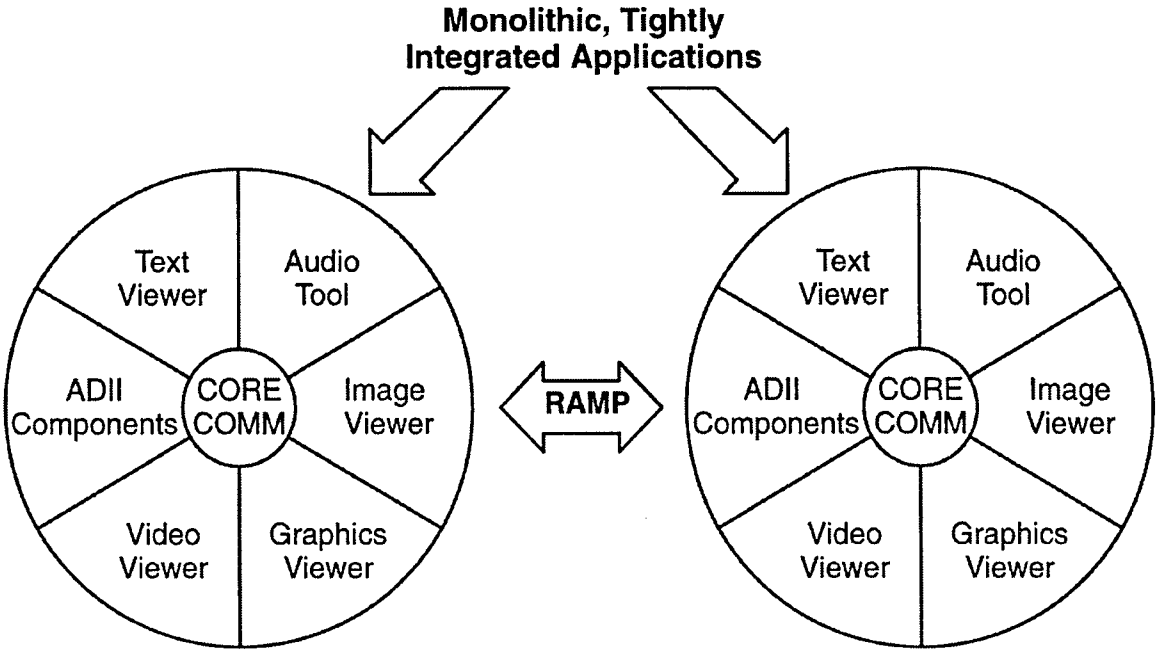


Figure 4-1 Original ImNet CSCW Architecture

IPC protocol was developed cognizant of InSoft's OpenDVE protocol. With minimal effort, our modules could be plugged in to InSoft's Communiqué! or similar conferencing system, greatly facilitating code reuse.*

4.2 Conferencing Scheduling and Initiation

After the core CSCW component was enhanced, we incorporated the Conference Scheduling and Initiation System (CSIS), also developed under ImNet, into the VIEP testbed. CSIS supports the creation of collaborative conferences.

When the user first logs in to a workstation, he launches the CSIS Registration Interface, shown in Figure 4-3. A simple startup file in the user's home directory specifies his name and additional parameters to the CSIS. The user's name is shown in the upper right of the window. Below that, there is a list of active conferences. Any conferences that the user has been invited to join will be shown here. As shown, this user has one active conference called "Sarajevo Scenario." The application that will be launched when this conference is joined is shown in parentheses after the con-

*Since the initial technology survey, InSoft has ported their Communiqué! and OpenDVE software to SGI machines. Although we could plug our modules into Communiqué! with little effort, their per-seat cost of approximately $2000 makes this cost prohibitive at this time. In addition, our own CSCW infrastructure now incorporates all functions found in Communiqué that are needed for the C^3I scenarios that we have been investigating.

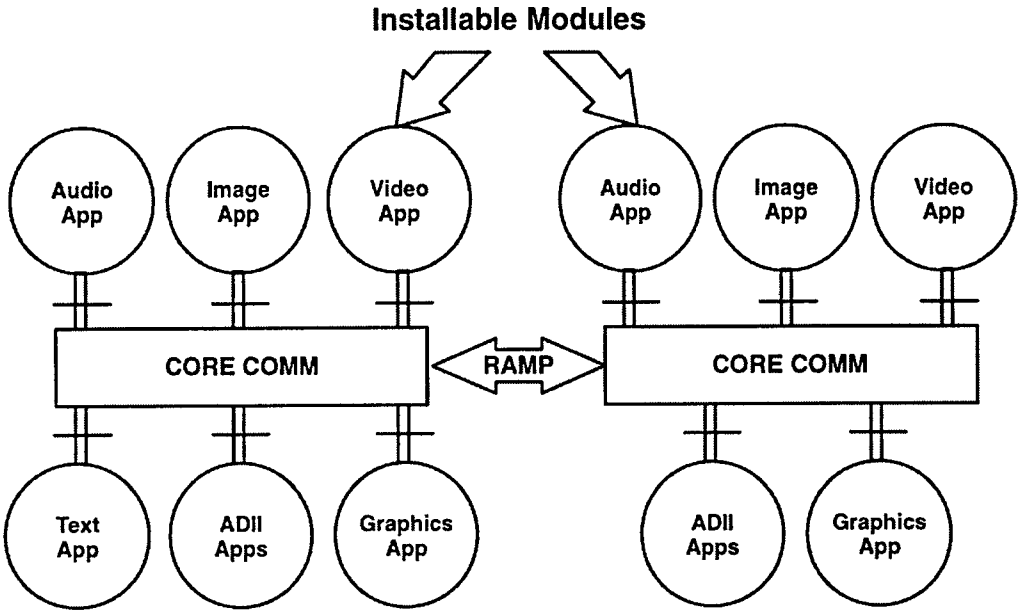


Figure 4-2 New VIEP CSCW Architecture

ference name. In this case "CSCW@TASC" will be launched. This particular application contains the collaborative components of the VIEP testbed.

A list of currently-scheduled conferences is obtained by selecting the *Schedule* button,

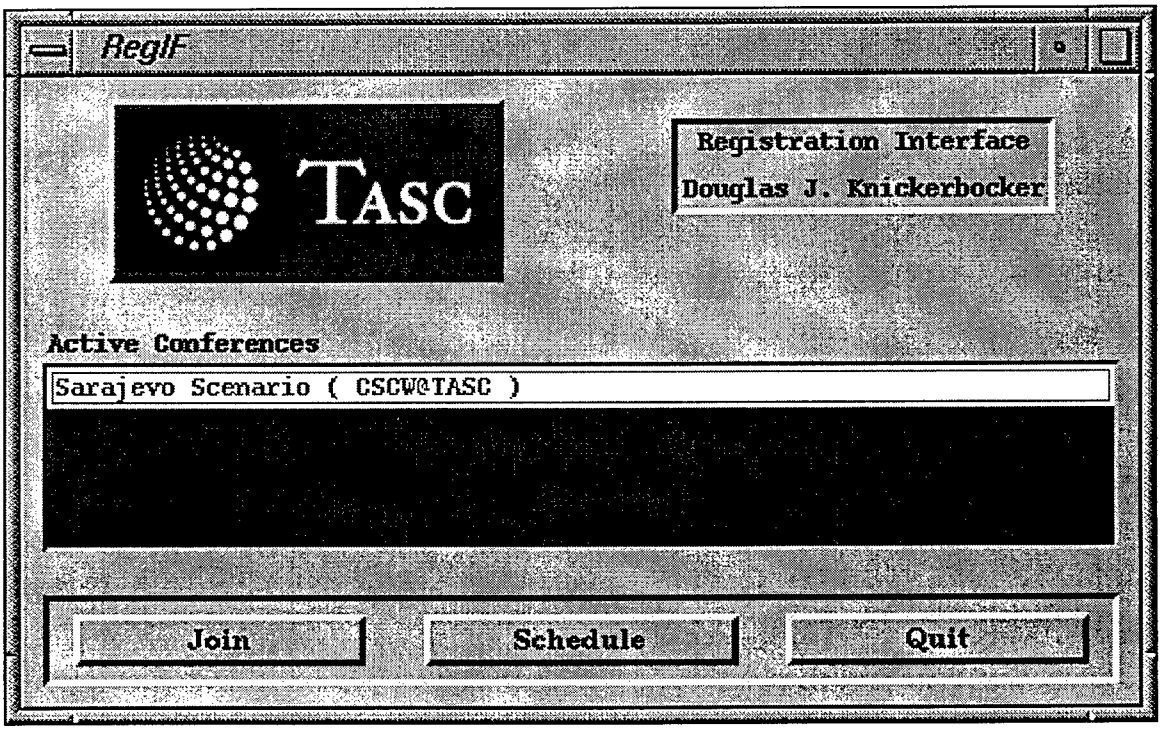


Figure 4-3 CSIS Registration Interface

which pops up in the CSIS conference listing window shown in Figure 4-4. The set of conferences currently managed by the CSIS are shown in the main portion of the window, along with an indication of whether the conference is in progress or scheduled for future execution, the date and time of the conference, the conference title, which application the conference will launch, and who created the conference. If a conference is selected, the participants (besides the creator) are shown in the "Participants" section of the window. Conferences can only be modified or deleted by the conference creator.

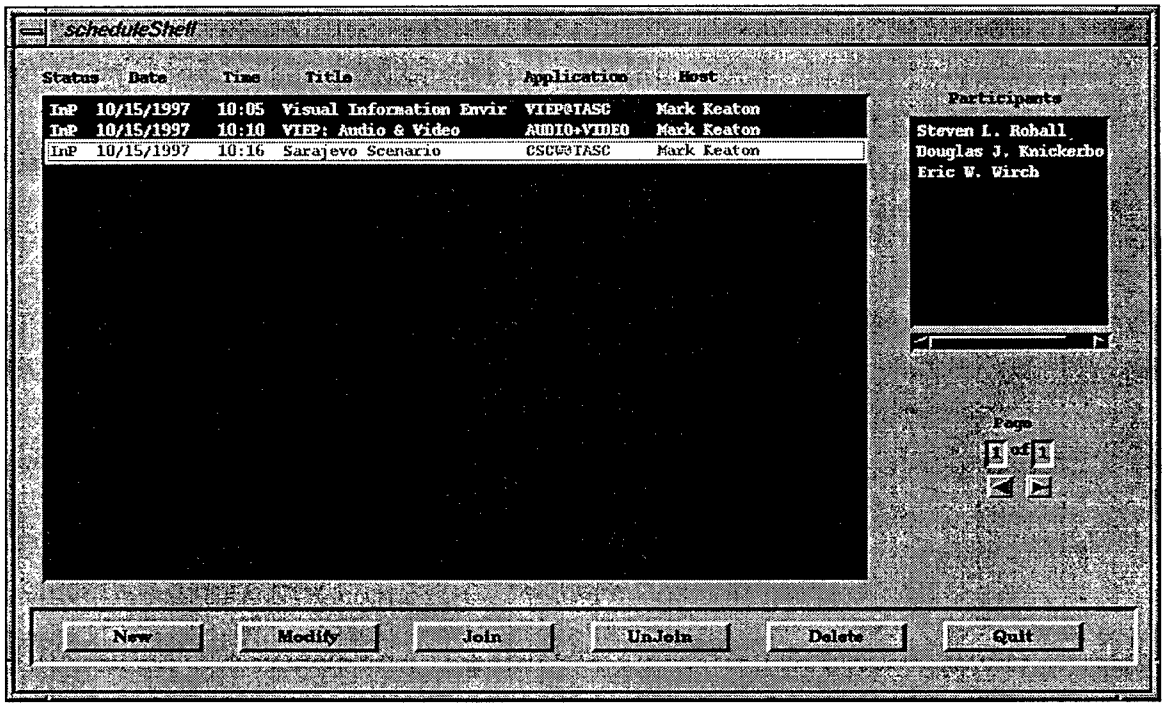


Figure 4-4 CSIS Conference Listing

A new conference is created by selecting the *New* button, at which point the conference creation window shown in Figure 4-5 appears. The conference creation window allows users to name conferences as well as to specify their starting times. The calendar and clock tools allow users to specify start times as the current time or any time in the future. The default time is "now," i.e., a conference will start as soon as it is created. Conference creators also specify the set of participants that should be included in the conference. Conferees are specified by name; CSIS takes care of finding participants regardless of their network address. This is especially useful for dial-up users with dynamically assigned IP addresses.

The final step in conference creation is specifying the application used for the collaborative

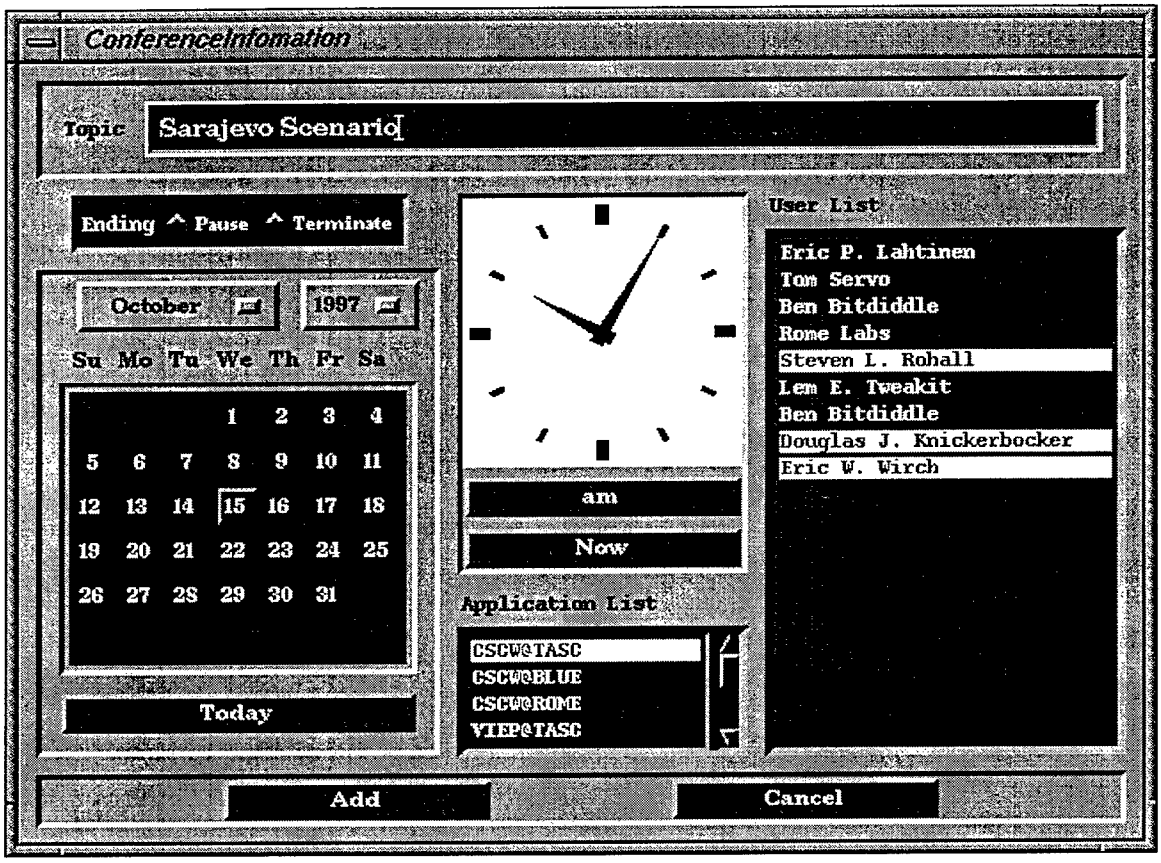


Figure 4-5 CSIS Conference Creation

session. The application is selected from the application list. Presently, there are two main applications used for the VIEP scenario: the applications titled starting with "CSCW@" (there are several variants depending upon where the demonstration is being given) which start TASC's CSCW conferencing system, and audio/video teleconferencing. (The audio/video teleconferencing application is shown in Figure 4-6; the remainder of this document will describe TASC's CSCW conferencing system in more detail.) As new applications become available, they can simply be "plugged-in" to the CSIS environment without requiring any code changes to CSIS itself.

When a conference is launched, all conferees are automatically notified (by an audible signal) of their invitation to join. CSIS retains any unanswered invitations even after a conference is launched, and will notify any new participants immediately upon log in. When a user joins a conference, all conference applications are automatically launched for that user.

4.3 The TASC CSCW System

When a user joins a collaborative VIEP conference, he is first presented with the CSCW

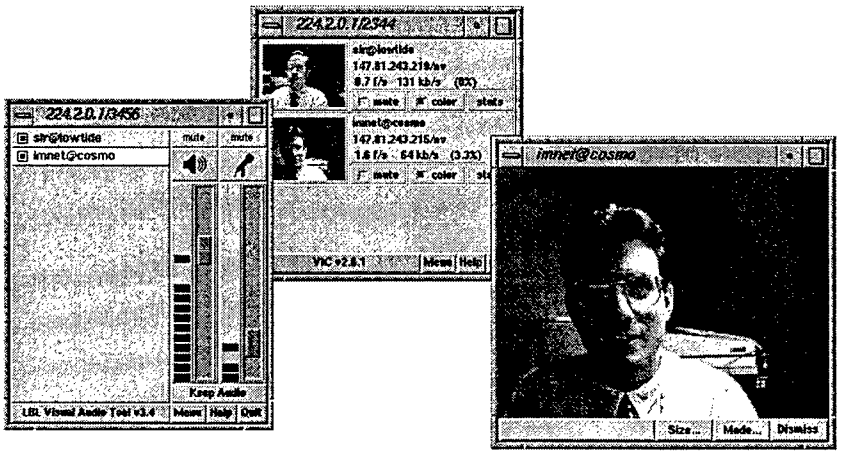


Figure 4-6 Audio and Video Teleconferencing

desktop (Figure 4-7). The desktop contains all components for the collaborative session including:

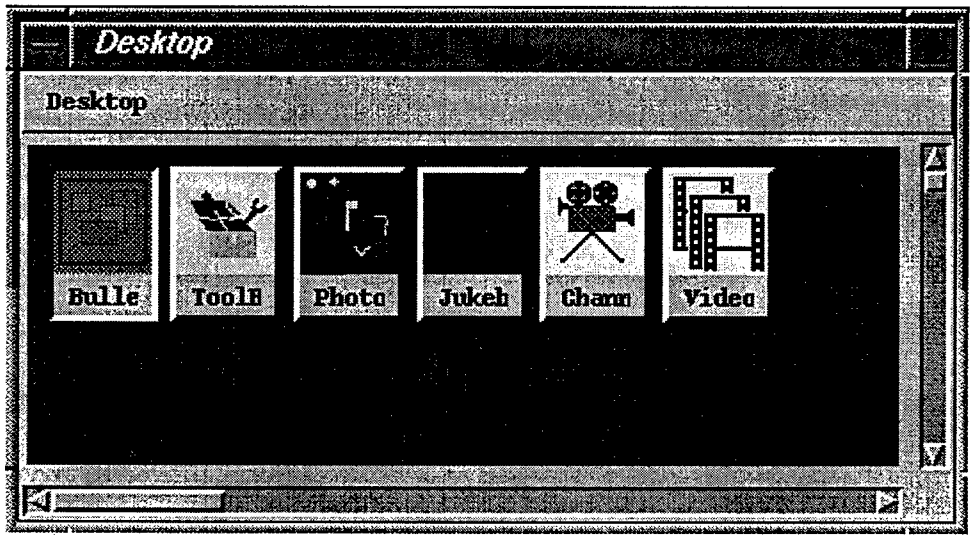


Figure 4-7 CSCW Desktop

Bulletin Board - The central item on the desktop is the conference bulletin board (Figure 4-8), which displays the shared state of the conference. Everyone in the conference has access to everything that appears on the bulletin board. In this example are shown icons representing the collaborators in the conference, several multimedia information sources (an image "sara," an audio clip "CNN," and an MPEG movie "sara"), as well as several collaborative viewing tools (an image viewer "Image," an audio player "XAudio," and an MPEG movie player "Movie").

Associations are made between viewing tools and data sources by dragging and dropping

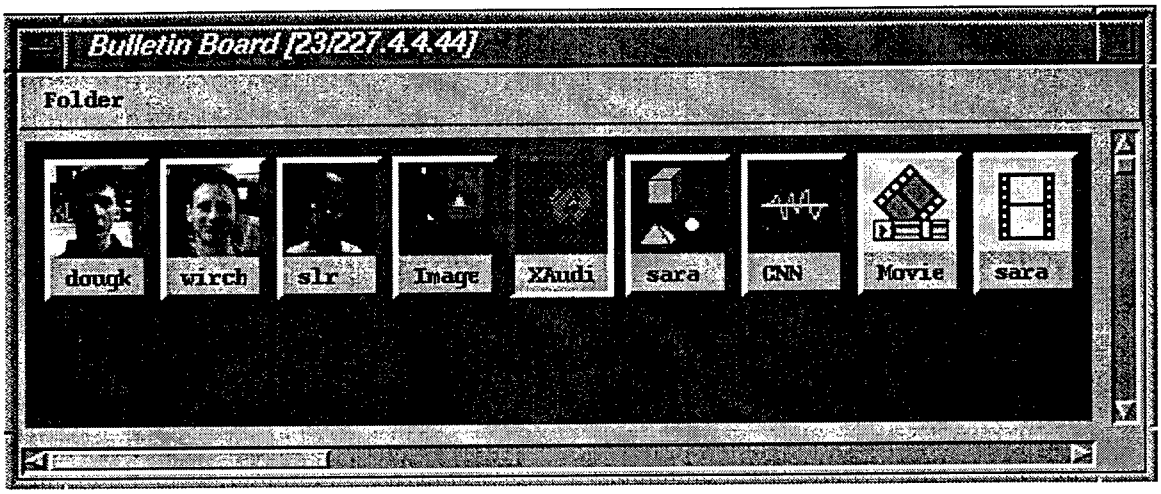


Figure 4-8 CSCW Bulletin Board

the respective icons. For example, an audio clip can be associated with an audio player; when the audio player is double-clicked, an audio player is launched that will enable all users on the bulletin board to hear the audio

Tool Box - The tool box (Figure 4-9) contains multi-user applications for use within the collaborative session. At this time, the VIEP testbed has five such tools: two high-performance collaborative image viewers (one augmented with speech recognition capabilities), a collaborative audio player, a collaborative MPEG movie player, and the DIVA MPEG movie player capable of viewing streaming video. These tools will be shown and described in more detail below.

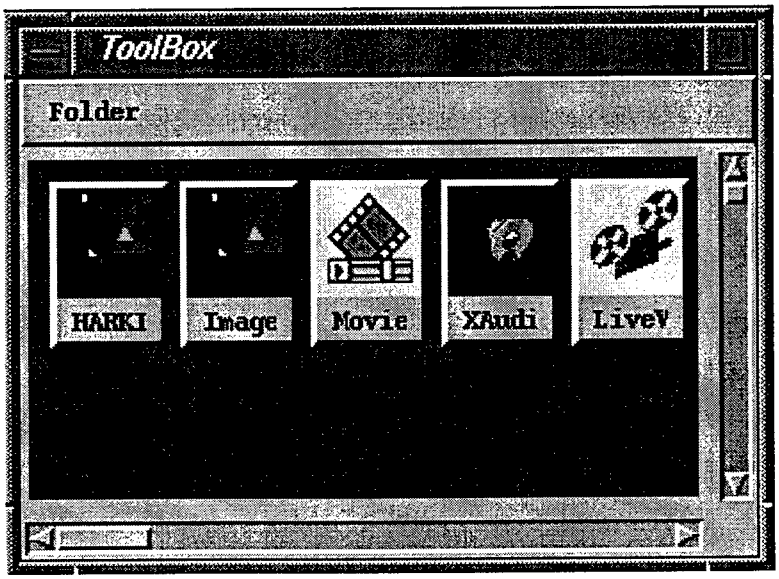


Figure 4-9 CSCW Tool Box

Photo Album - The photo album (Figure 4-10) contains icons representing various images that may be used within the conference.

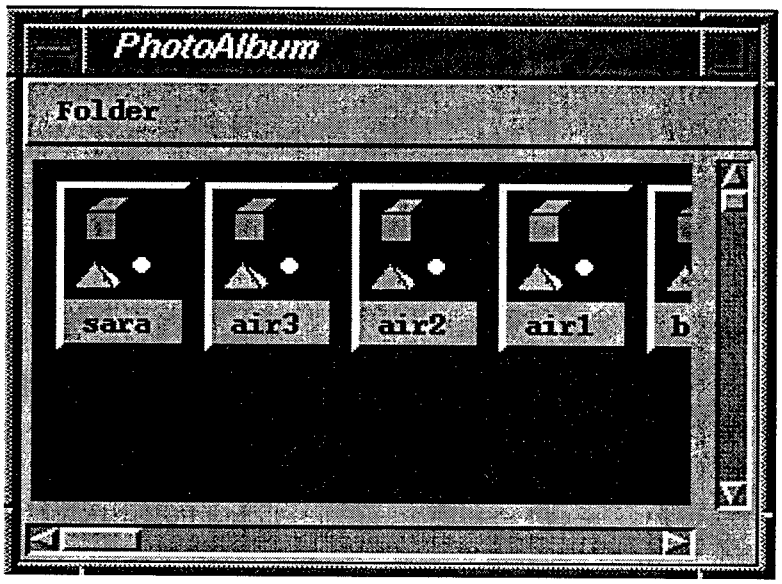


Figure 4-10 CSCW Photo Album

Jukebox - Like the photo album, the jukebox (Figure 4-11) contains various recorded sounds that may be used within the conference.

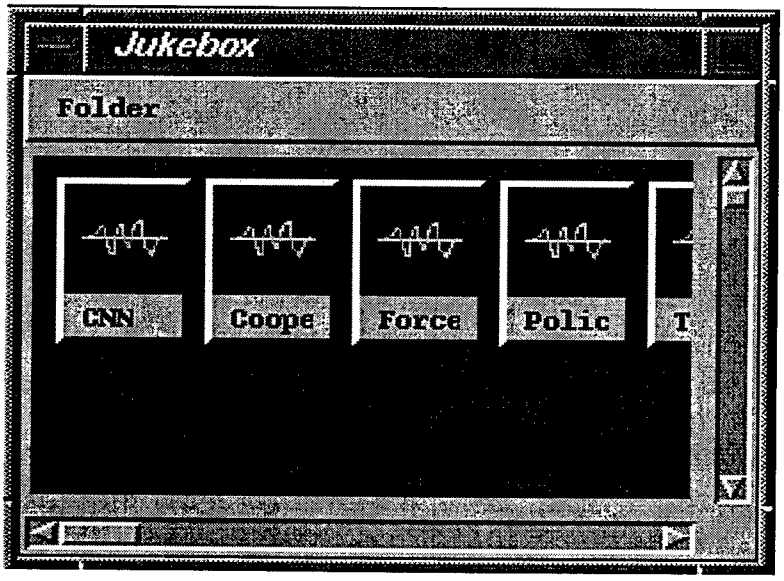


Figure 4-11 CSCW Jukebox

Channels - Like the photo album and jukebox, the channels folder (Figure 4-12) contains

MPEG 1 and MPEG 2 movies and servers for streaming MPEG video that can be used within the conference.

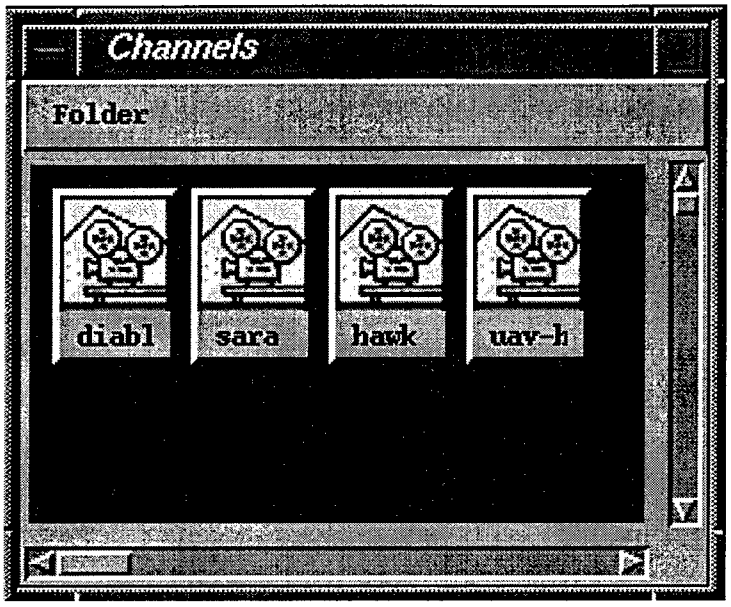


Figure 4-12 CSCW Channels

Video Store - Like channels, the video store (Figure 4-13), contains MPEG 1 movies that may be used within the conference.

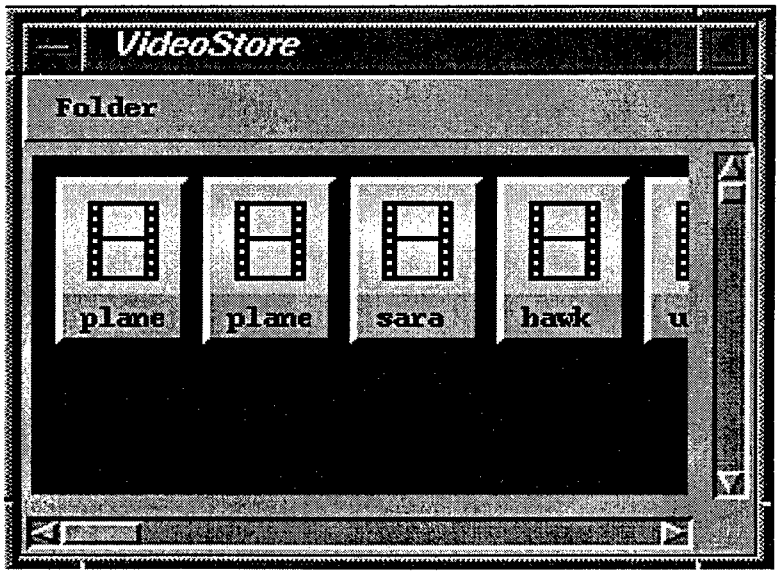


Figure 4-13 CSCW Video Store

Double-clicking any of the icons on the desktop will open the respective window. These

windows can be opened and closed as needed. The state of the windows and the actual position of any of the icons may vary from user to user. Only the state of the bulletin board is shared among users. Even with the bulletin board, different users may have the window open or closed or have it sized differently or have the icons placed differently. This information is not shared among the users.

4.4 Collaborative Tools

TASC has developed several unique collaborative tools for use within the VIEP demonstration. These include an image viewer capable of viewing images of practically any size, an audio player, an MPEG 1 video playback module, and the DIVA video playback module capable of viewing streaming video.

4.4.1 Image Viewer

The high-performance collaborative image viewer in the VIEP testbed (Figure 4-14) was originally developed under the U.S. Government-sponsored IMACTS program. The viewer is able to handle images of extremely large size when they are stored in the AIMS image format (a TIFF-based, tiled image format originally developed under the ImNet program). AIMS images are organized as multi-resolution pyramids. The viewer shows the user a low-resolution overview of the image and allows the user to zoom into and out of areas of interest. At any particular resolution, only the tiles needed to fill the high-resolution portion of the viewer are fetched. Additional tiles are fetched as the user scrolls and zooms the image. The user can also adjust the contrast and brightness of the image.

Under VIEP and TASC internal funding, the viewer was made collaborative. As described above, an image is associated with the viewer. When the image viewer icon is selected, the viewer will be launched on the workstations of all the conferees. Image loading, panning, scrolling, and zooming functions are then shared among these viewers.

Telepointers are also supported within the collaborative viewer. The viewer allows each user to control an independent telepointer; any number of telepointers can be shown and manipulated simultaneously. To aid interpretation of a potentially large number of telepointers, each user can independently assign a color to his pointer. When a user presses a mouse button in the high-resolution portion of the viewer, his pointer location is shown in all of the viewers in the specified color. In Figure 4-14, a remote user's telepointer can be seen as a small yellow square near the plume of smoke in the center of the high-resolution image.

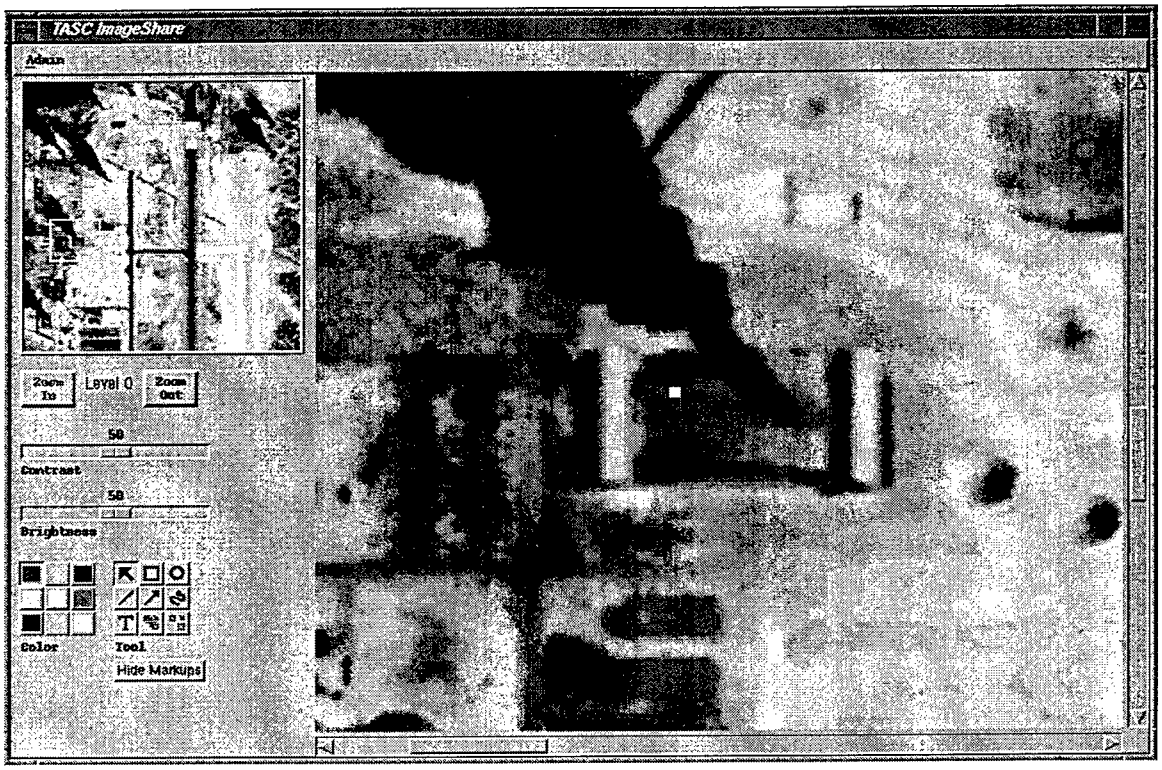


Figure 4-14 Collaborative Image Viewer

Collaborative annotations have also been added to the image viewer. Annotations, including squares, circles, lines, arrows, free-hand sketches, and text provide for enhanced understanding of the image artifacts during collaboration. Collaborators can create annotations simultaneously and, in fact, all users can see annotations as they are being created. For example, all users see a free-hand sketch as the user creating it moves his mouse in the high-resolution portion of the viewer. The annotations are drawn in both the high-resolution portion of the viewer and in the respective location of the low-resolution overview portion of the viewer. Users can choose different colors for their annotations and are also allowed to move or delete the last annotation created, even if they did not create it. Figure 4-15 shows the image viewer with several annotations.

To demonstrate the utility of collaboration to the task of mission planning and real-time replanning, a collaborative mission planning tool has been developed. At the heart of this tool is an iconic overlay capability. The icons are a specialized form of image annotation used to represent various battlefield components. The icons may be animated; these moveable icons allow the image viewer to be used for mission monitoring. The user interface for the collaborative mission planning tool is shown in Figure 4-16.

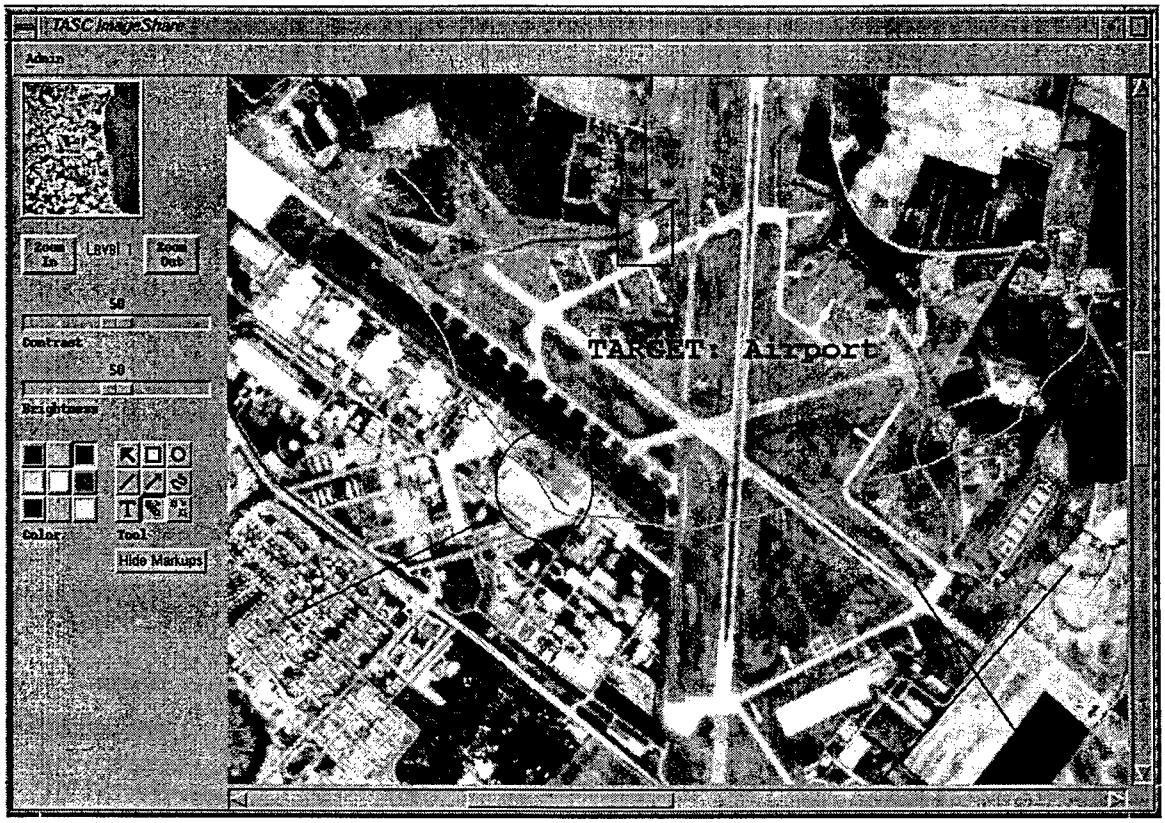


Figure 4-15 Image Viewer Annotations

Figure 4-16 shows an image of the area around the Adriatic Sea in the image viewer. Overlaid on top of the image are icons for several battlefield elements including a carrier, an A10, two F-15, an two F-18s. Also shown are the flight paths for each of the elements. The yellow portion of the flight path indicates when a plane is en route to a target, red indicates that the plane is over the target, and white indicates that the plane is leaving the target area.

New flight paths are created using the pop-up window shown in the upper left of Figure 4-16. The user selects a button on the lower right of this window for the type of battle field element to add to the scenario. He then draws the flight path free hand using the mouse in the image viewer window. Clicking the mouse button while drawing the flight path indicates to the system when the element is over the target area.

As new flight paths are created, they are also shown on the time line on the mission profile indicator portion of the mission planning tool. A label on the left of the window provides the name of the battlefield element. The bar to the right of the label shows the movement of that element during the time of the mission. Again, yellow represents the element on the way to a target, red is over

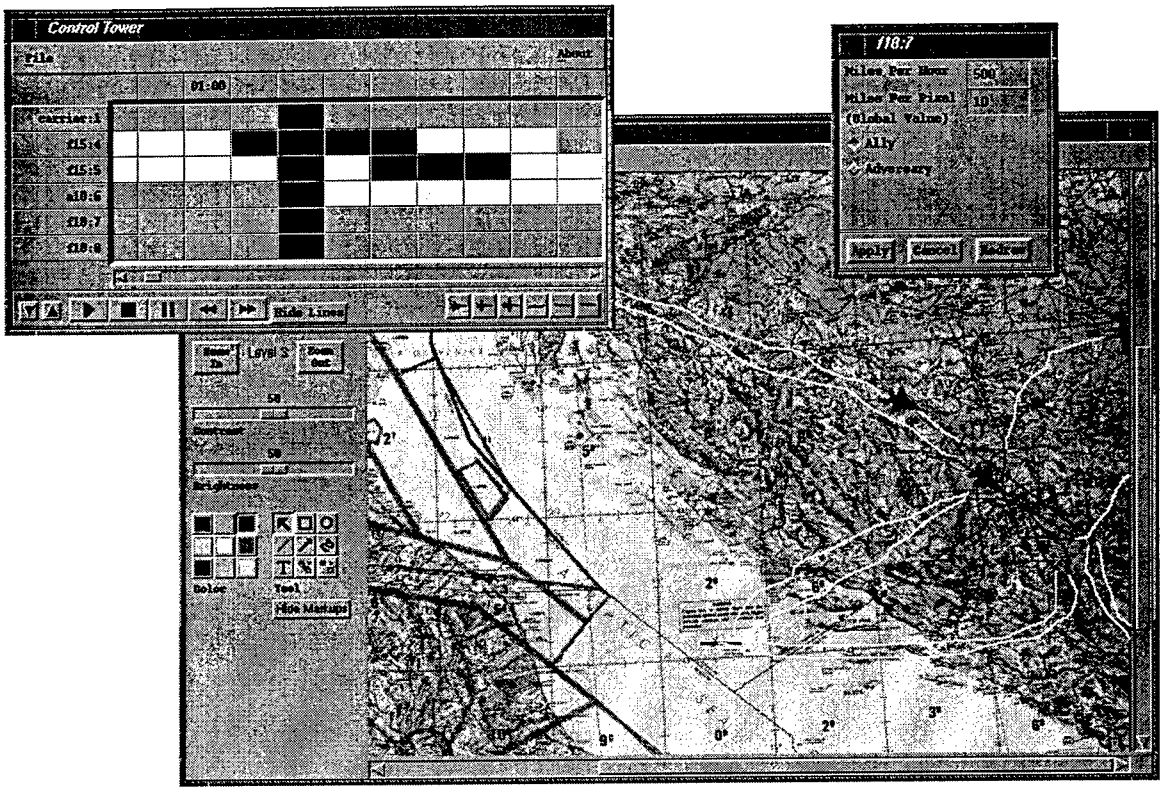


Figure 4-16 Collaborative Mission Planning Tool

the target, and white is leaving a target. The black vertical bar on the time line shows the current time in the mission scenario.

The other buttons on the pop-up window control the playback of a scenario. The user can "hide lines," meaning that the actual flight path traces should not be drawn on the image viewer. There are VCR-like controls for playing and pausing the scenario playback. Finally, there are buttons for controlling the playback speed; the scenario can be played back either slower than real time (e.g., for training purposes) or faster than real time (e.g., for "what-if" planning during an exercise). Playback is collaborative; any user can start or stop the scenario and it will start or stop for all users. The mission tool utilizes RAMP for efficiently multicasting the playback commands to all VIEP users.

Battlefield elements can be edited by selecting the appropriate label on the left of the window. A user is able to modify each battlefield elements' speed and choose if the element is an ally or and advisory. Also, redrawing the flight path of each battlefield element is possible by selecting the redraw button.

Missions can be saved and loaded from an option on the "File" menu. The file format is ASCII text that can be created from the output of other systems or massaged into the appropriate format as input to other systems. This capability is useful for creating the beginning of a mission and then modifying and augmenting the scenario as needed as new intelligence data is received.

Although this mission playback capability is fairly limited, it demonstrates how VIEP can be used for mission planning and real-time replanning. As intelligence data is received from the field, commanders are able to analyze the situation in consultation with others, to investigate several options for modifying the mission, and to order operational changes, all while a mission is in progress.

4.4.2 Audio Player

A multi-user audio player has also been created for VIEP. As with the other collaborative tools, an audio clip is associated with the player. When the player is selected, the audio clip will play at all users' workstations. The user interface for the player, shown in Figure 4-17, has VCR-like controls for controlling audio playback. Any of the users can cause the audio to be played or paused or can associate a new audio clip with the player. In addition, a scrollbar is provided for allowing random access to the audio clip; a user can grab the scrollbar and drag it to listen to any portion of the audio clip. The time counter shows which part of the audio is currently playing. Each user also has independent control, with the up and down arrows, over the volume of his local playback.

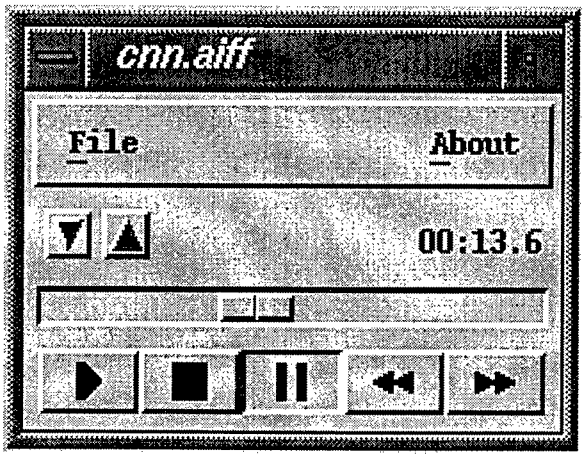


Figure 4-17 Collaborative Audio Player

Presently, the audio clips are stored at each user's location to provide enhanced response time. Streaming audio (similar to VAT sessions) for larger stored audio clips or real-time feeds

could be supported at the expense of much greater set-up time and network utilization.

4.4.3 MPEG 1 Video Player

An improved multi-user MPEG 1 video player (Figure 4-18) has been integrated into VIEP. Like the audio player, the video player decodes (in software) and plays color MPEG 1 videos stored at each user's workstation. The new player has improved VCR-like controls that allow the users to play (forward and reverse), freeze frame, rewind, and single-step the video (forward and reverse). Users can also seek to a particular frame, loop the playback, and play in fast forward. All of these operations are synchronized among all viewers. The new MPEG player also has faster decoding and display.

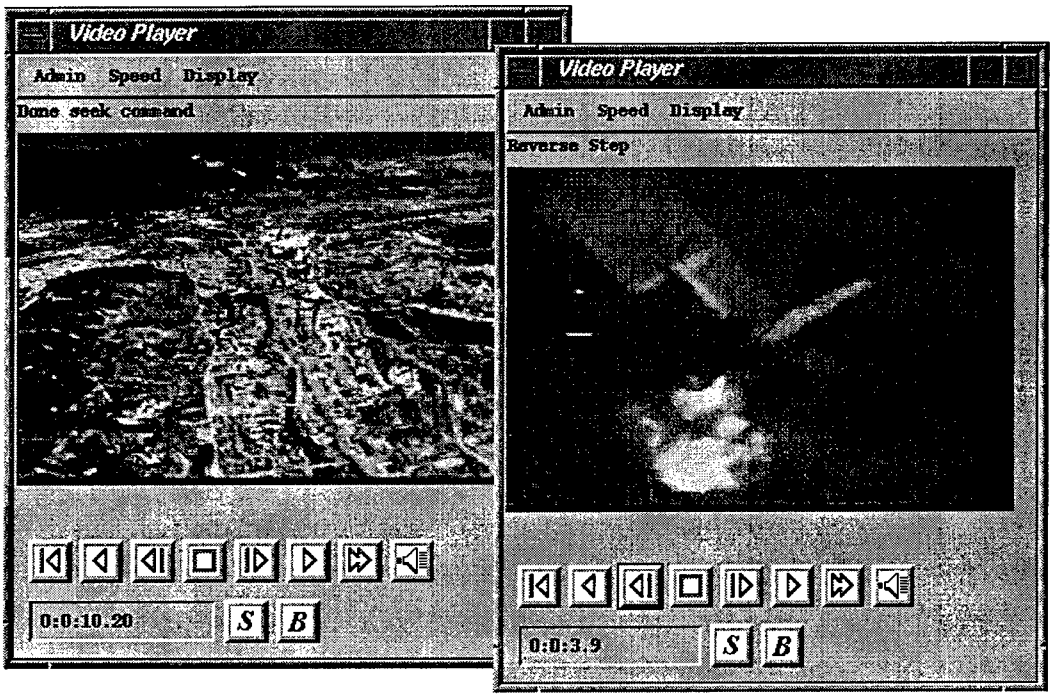


Figure 4-18 Collaborative MPEG1 Player

4.4.4 DIVA Streaming Video Player

A second video player (Figure 4-19) has been integrated into VIEP. Originally developed under the Rome Laboratory sponsored DIVA project, this new video player is capable of viewing streaming video from a server running an MPEG 2 encoder. The server multicasts (point-to-multipoint) the video stream for efficiency. In addition to the capabilities of the MPEG 1video player, the DIVA video player also has a zoom feature, which allows high-resolution playback, that is shared among each of its users. Another feature, which is not shared, allows users to record video

that is being streamed from a server.

Figure 4-19 DIVA Streaming VIdeo Player

4.4.5 Incorporating Third-Party Applications

As described above, the VIEP collaborative infrastructure is ultimately extensible since we control the source code and have defined an API for incorporating new applications. However, adding a new application to VIEP requires that it be recompiled in order to include the appropriate API calls. When an application cannot be recompiled, it may be still be used collaboratively by sharing its outputs.

To demonstrate this, we incorporated the MUSE pscene application into the VIEP demonstration. MUSE is a software package available from NIMA for image manipulation. pscene is used to generate 2-1/2D perspective renderings using DTED data in conjunction with georegistered ARDG raster maps (Figure 4-20). Although we could not modify the pscene application and therefore could not collaboratively generate the perspective scenes, we can share the output of pscene using the image viewer application.

With a simple program, we convert the saved pscene output to TIFF and create an icon for the new data item on the bulletin board. The TIFF image is compatible with our image viewer. An image server is then used to multicast the data to the remote collaborators since they may not have direct access to the file with the image data. With the ability to efficiently serve imagery to remote users, we can share the output of what are otherwise single-user applications.

Figure 4-20 MUSE Software Package

4.5 Wireless Interface Capabilities

Natural, untethered interactions with the system are deemed critical for acceptance by high-level commanders. We have incorporated speech recognition, video-based gesture recognition, and wireless laser pointers and text entry into VIEP as first steps towards achieving this goal.

4.5.1 Speech Recognition

Initially BBN's HARK system was integrated into the image viewer to provide a natural language interface. HARK is speaker independent and recognizes continuous speech (phrases). The panning, scrolling, and zooming functions of the collaborative image viewer are all accessible via voice input. First a grammar was developed for understanding commands which might be used in the control of an image viewing tool. By speaking commands such as "scroll north," "go south," or "zoom in" the user is able to control image position and magnification. By repeating a direction the user increases the scrolling speed in that particular direction. To maximize flexibility, the system has been configured to accept both spoken language controls and traditional mouse controls. For example, a user can use the mouse to zoom in, then use voice control to move to an appropriate section of the image. A user controlling the VIEP system using voice commands can be seen in

Figure 4-21. The monitors behind the user show the collaborative image viewer synchronized between two workstations.

Figure 4-21 Speech Control of the VIEP System

One problem encountered with HARK development was the desire to control more than one application at a time with HARK. Typically, this is accomplished by forking separate HARK processes, one to handle each application. However, HARK processes consume a lot of CPU resources and using more than one process unacceptably slows the workstation. To solve this problem, we implemented a HARK server capability for use within VIEP.

The HARK server runs on any machine in the network that has a HARK license. One grammar file is created to define the grammars for all clients. Client applications (on possibly separate machines) connect to the HARK server. When speech is recognized, a "speech recognition event" is sent to the client much the same way that a key press event is sent to an X Window System client. Our client applications were modified to connect to the HARK server if an environment variable was defined and to fork a new HARK process if the variable is not defined.

4.5.2 Video-based Gesture Recognition

The second generation of Sandy Pentland's gesture recognition software has also been integrated as a second wireless interface component. The software detects motion by subtracting a

pre-acquired background image from the incoming video stream. The difference image is fed to an image understanding module that infers the location of various body components, allowing hand positions to be isolated. Using standard SGI equipment, the system is capable of tracking whole-body motion at approximately 15 frames per second with a resolution of 160 by 120 which is adequate for initial demonstrations.

The gesture recognition system runs as a server; any application can connect to the server in order to get gesture events. Initially, the gesture recognition software has been used to control a telepointer on the shared image viewer. When the user moves his hand, it is as if he is using the mouse to telepoint to the remote collaborators. The user sees visual feedback on the image viewer of where he is pointing; this location is then transmitted to all conferees.

Figure 4-22 shows the gesture recognition system in use. The user is standing in front of an SGI workstation which has an IndyCam connected to it. The window titled "backsub" shows the output of the image understanding module; this is what the system "sees" after the pre-acquired background image is subtracted from the live video stream. Ultimately, the hand position is used to control a telepointer on the collaborative image viewer. In Figure 4-22, the telepointer is the small yellow square below the plume of smoke in the center of the image.

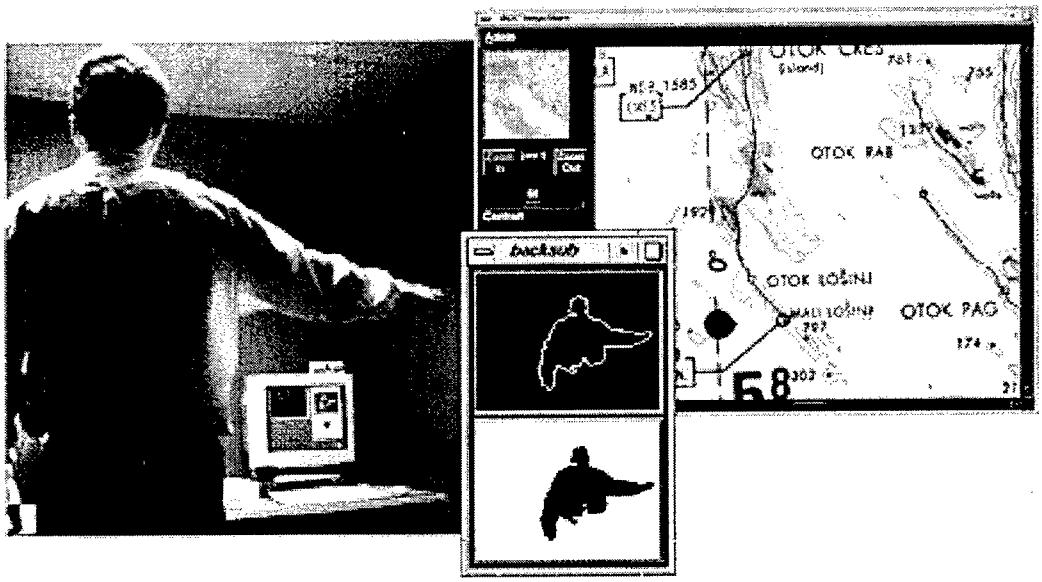


Figure 4-22 Gesture Control of the VIEP System

While the current gesture recognition system shows great potential, it does have several shortcomings. First, the system is limited to tracking whole body motion. This provides for very coarse gesturing control. As shown in Figure 4-22, the system is capable of detecting when a user

has extended his arm away from his body. However, if the user moves his arm so that it is in front of his body, the system can no longer see it; there is no way to determine if the arm is in front of the body or if it is straight down along the side of the body. This problem is particularly troublesome since people naturally tend to point things out by raising their arm and pointing in front of their body.

The second shortcoming is that the gesture recognition system is only suitable for standoff operation. As the user approaches the camera (which would in practice be mounted above a large screen projection system) he either obscures the background, thereby confusing the background subtraction module, or he leaves the camera's field of view totally. Improved software now in development could address the first shortcoming. A different form of wireless, near-screen pointing capability is needed to solve the second problem.

4.5.3 Laser Pointers

As discussed earlier, we have been investigating the use of inexpensive laser presentation pointers as an electronic "grease pencils." A camera behind the screen of a rear-projection system could be used to detect a laser spot on the screen. That information could then be used to control the system in place of a traditional mouse. Although the problem sounds straight-forward, implementing a system that works well in practice is non-trivial.

Rome Laboratory personnel implemented a first version of a laser tracking system. Initially, a PC equipped with a video input card was used for the image processing. A second version of this system was ported to the SGI platform utilizing the SGI Sirius video board. TASC received the software from Rome Laboratory and initiated a port to the SGI VINO video board. The rear projection display system installed in the VIEP laboratory at TASC can be seen in Figure 4-23. Note the laser pointer spot on the screen near the middle of the map within the collaborative image viewer.*

Several problems were encountered during this initial port that required several modifications to the code. All of the problems can be traced to reflections of the projector guns off of the back of the projection screen within the view of the detection camera. Basically, if the detection camera can see the reflections of the projector guns, then it cannot reliably detect the laser spot. The difference in brightness between the darkest spot of the projected image and the laser spot is

*A third generation system has been ported back to the PC platform for performance.

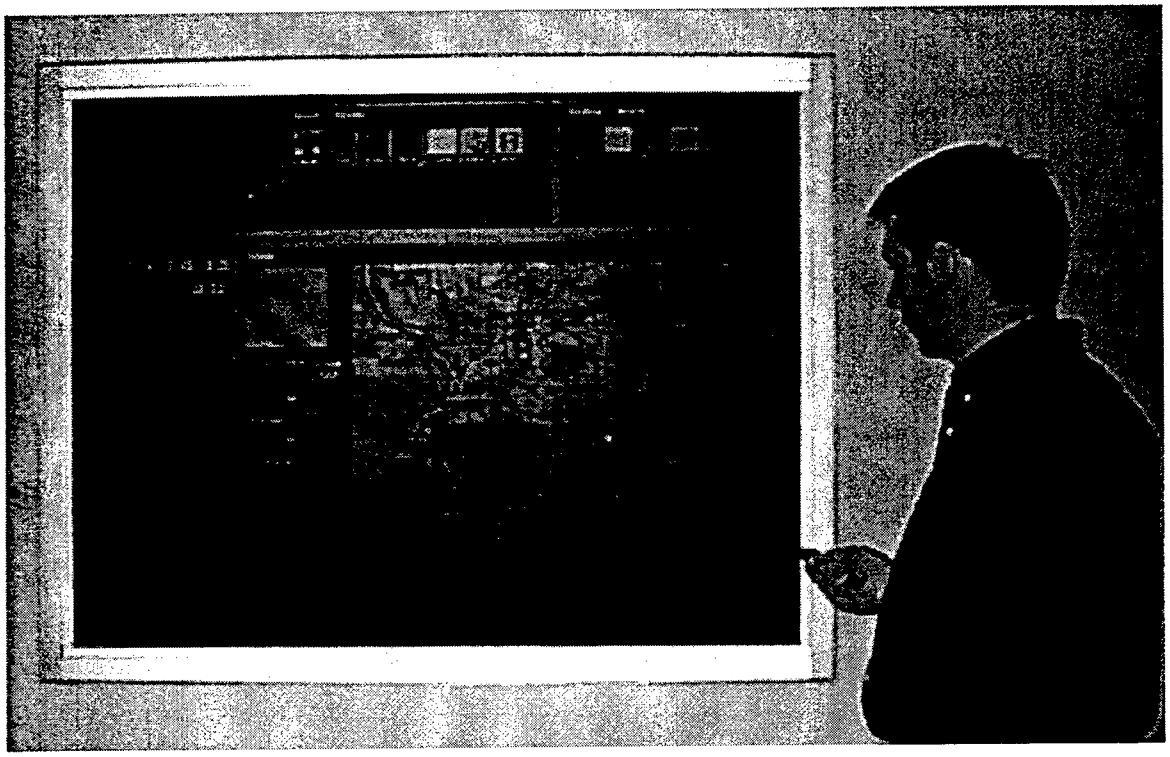


Figure 4-23 Rear Projection Display at TASC

not that large. The reflections of the projector guns appear much larger and much hotter than the laser spot. This problem can be seen in Figure 4-24. The reflections from the red, green, and blue projector guns can clearly be seen. The laser spot that the detection system is trying to find is the small, bright spot down and to the left of the gun reflections.

There are several approaches to addressing the reflection problem. Perhaps the most straightforward is to eliminate the source of the reflections--the projection screen glass. Instead of glass, fabric can be used to provide the screen surface. The rear of the fabric is not reflective enough to cause the projector guns to be seen. As a result, the detection camera can be placed directly behind the screen thereby simplifying the translation of the detected laser spot into coordinates on the X Window System display. This approach was ultimately adopted at the facility at Rome Laboratory. However, removal of the glass projection screen was not an option for our laboratory at TASC.

A second approach we tried in addressing the problem involved the use of an optical band pass filter. Inexpensive laser presentation pointers generate light at a wavelength of 670nm. We were able to purchase a bandpass filter centered at 671nm with 50% transmission at ±2nm of the center. This filter initially showed great promise; while one could not see through the filter, the la-

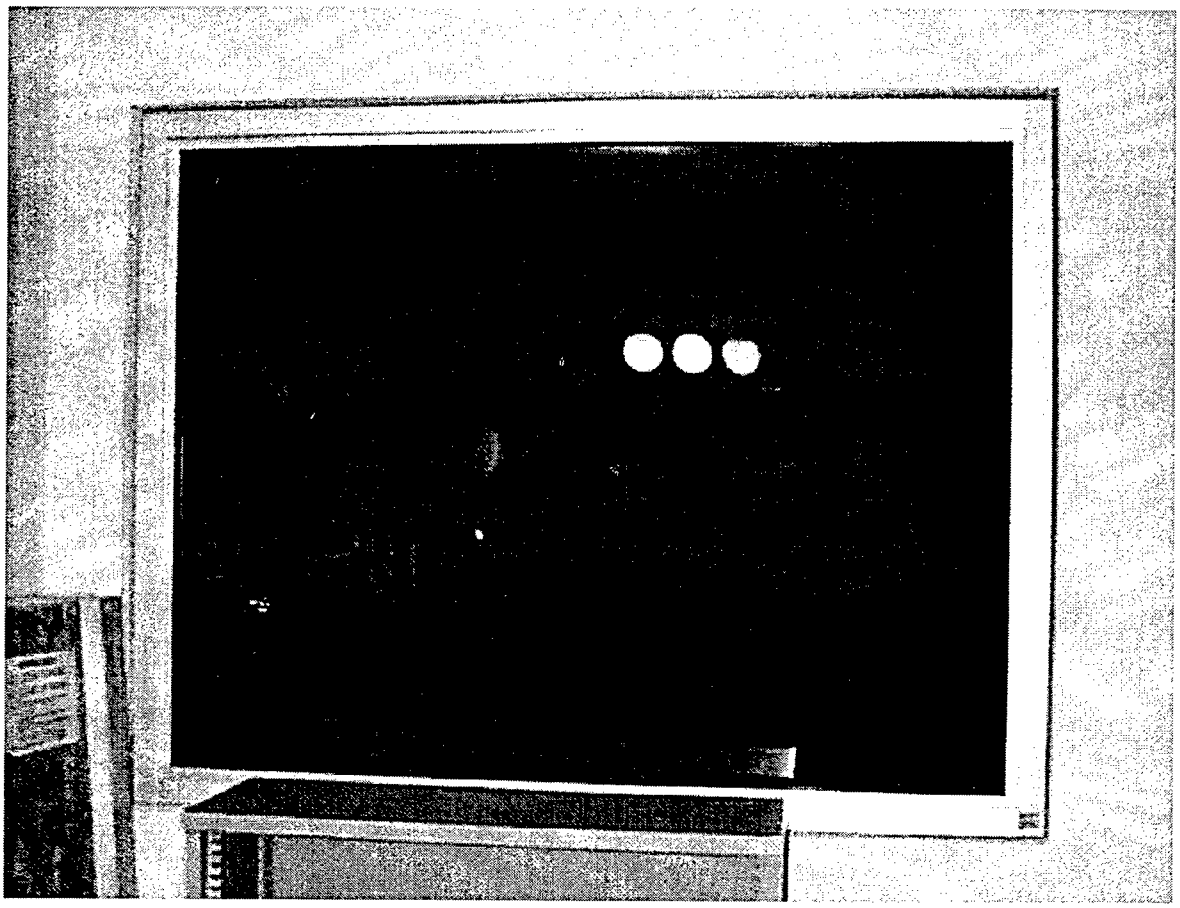


Figure 4-24 Reflections From the Projection Screen

ser pointer passed through as if it were clear glass. Our hope was that the filter would allow us to keep our detection camera directly behind the projection screen thereby simplifying the spot translation problem. However, the filter introduced new problems. In particular, the filter greatly diminished the brightness of the laser spot seen on the rear of the projection screen. This problem could be addressed, though, since other visible transmissions from the projection screen were reduced to nearly nothing. The bigger problem was that the thickness of the filter was such that, although it did a good job of allowing transmission of the laser spot when it was centered on the projection screen, it reduced the brightness of the spot around the edges of the screen to the point where it could not be reliably detected. Thinner filter media is not readily available.

The approach we adopted at the TASC VIEP facility is to place the detection camera off to the side of the projection screen. The placement of our detection camera in relation to the projection screen can be seen in Figure 4-25. While this eliminated the reflection problem it introduced complications into the translation of the detected laser spot into coordinates on the workstation display. Further refinements in the software and the use of different filters, after consultation with

Rome Laboratory, allowed us to place the camera directly behind the screen with acceptable performance. The algorithms for handling off-angle camera positions, however, are still useful since it is very difficult to precisely align the camera with the projected screen image.

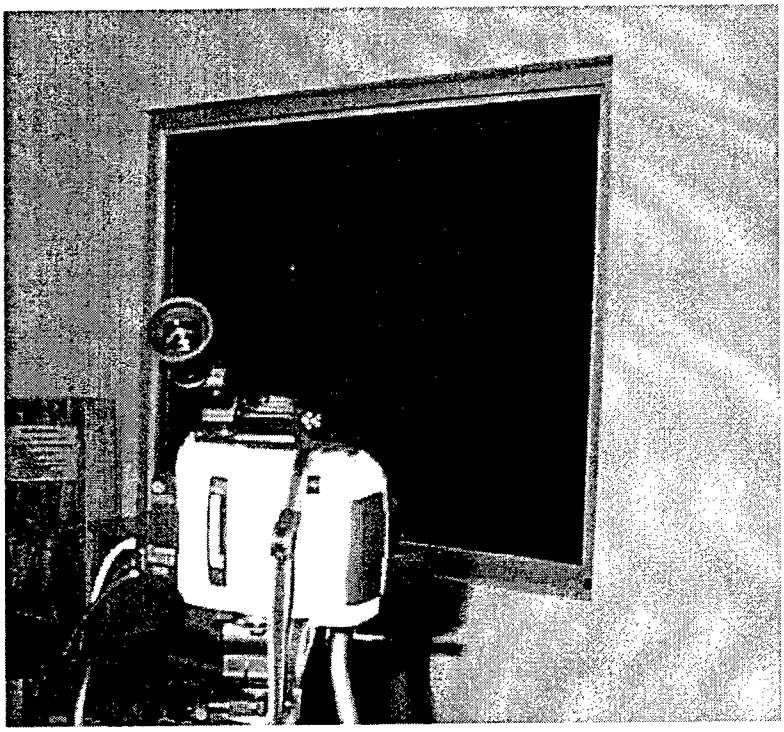


Figure 4-25 Camera Placement Behind Screen

When the detection camera is located directly behind the projection screen, the translation of laser spot location to coordinates on the workstation display is straightforward. There is essentially a linear transformation from laser spot to workstation coordinate. The original code received from Rome Laboratory required the user to indicate the upper left and lower right corners of the projection screen with the laser pointer in order to calculate the linear transformation between coordinate systems. When first tried at our facility, the tracking accuracy of the system was off considerably due to the fact that, when the camera is off-angle, the translation between coordinate systems is no longer linear. This can be seen in Figure 4-26 which shows the image that the detection camera is capturing.

To accommodate our off-angle detection camera, we needed to develop a more elaborate laser pointer initialization routine. The initialization begins with the display of a white window that covers the entire workstation display. This is used to determine the brightest possible matrix of image pixel values that the camera will see. This matrix is subtracted from the captured images when

Figure 4-26 View of the Detection Camera

the laser pointer system is tracking the laser. Doing this enables the laser spot to be more easily detected. Once the brightness matrix is captured, the user is asked to point to the four corners of the projection screen. This is used to calibrate the image the camera is seeing with the workstation display. We tried to implement an automatic detection routine where alternating bands of black and white would enable the system to detect the screen corners, but the difference between the black and white bands was not great enough to reliably detect the corners of the screen.

The detection and translation of the laser spot proceeds as follows. First, the detection camera captures an image which is digitized into the computer's memory. For efficiency, the neighborhood of pixels around where the laser was last seen is searched first. If the laser is not detected within that neighborhood, a scan of the entire frame is attempted. When a pixel is found that exceeds a preset threshold, it is used as the starting point for a Gaussian estimation of the center of the laser spot. The system works out from that spot looking for the direction of the greatest rate of change in intensity. A linear approximation followed by an installation-specific transformation is then used to translate the detected location into a workstation screen coordinate. The mouse on the

X display is then warped to the proper location. Tracking accuracy is quite good considering the low resolution of the video capture.

The user interface for the laser tracking system has been modified from the original code as well. Several examples of the interface can be seen in Figure 4-27. The interface allows the user to control which mouse button is being simulated (left, middle, or right). The laser pen itself can be used to manipulate the interface. A button press action occurs when the user shines the laser onto the screen. Alternatively, spoken commands interpreted by HARK or manual key presses can be used to manipulate the user interface or cause a button press action to occur.

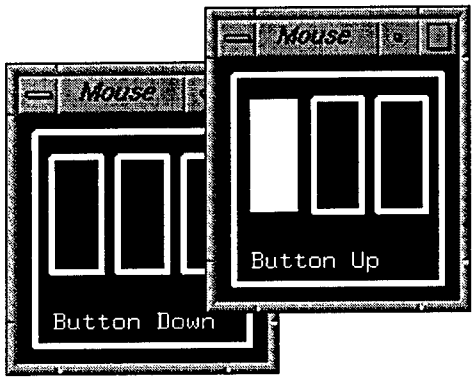


Figure 4-27 Laser Pen User Interface

4.5.4 XGraffiti

An application, XGraffiti (Figure 4-28), was designed specifically for use with the laser pointer and enables a user to enter text without using the keyboard. XGraffiti uses simple image processing based upon pre-stored templates to translate a shorthand-like input into an ASCII character. The select button allows the user to choose which window on the computer display will receive the corresponding key press events. A hidden keyboard enables the user to "hunt and peck" instead of using the shorthand notation.

4.6 Wide Area Networking

The WAN link between Rome Laboratory and TASC has been a dedicated T1 line. For the first and second year demonstrations, the T1 line was installed and routers integrated and configured to provide IP connectivity between the two sites. Providing the WAN link required resolution of several non-trivial technical and policy issues, including the design of adequate security firewalls and the establishment of the multicast tunnels required for RAMP operation.

For the third year and final demonstrations, we also used a dedicated T1 line. However, the

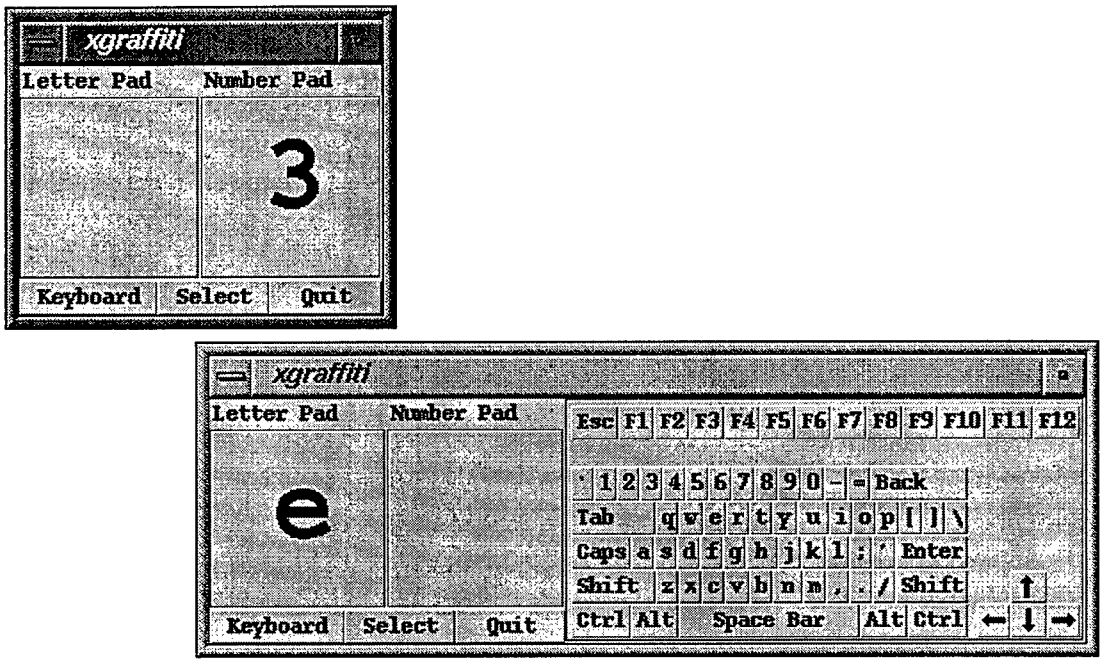


Figure 4-28 XGraffiti

end points of the T1 line were ATM switches provided by the Rome Lab-sponsored DIVA project (which is sharing demonstration equipment with VIEP). ATM interfaces in the demonstration machines were used to connect to the ATM switches. VIEP was running TCP/IP over the ATM network. This configuration provided a separate network that helped to isolate the VIEP demonstration from other network traffic at both TASC and Rome Laboratory.

A side benefit of having the network connection was its use for software installation and configuration on Rome Laboratory's machines. With the assistance of Rome Laboratory personnel, we were able to demonstrate the use of the VIEP system over the wide area network on several occasions.

5. SUMMARY

We have developed an example focus scenario centered around a real-time mission planning situation that demonstrates the need for collaborative capabilities and the utility of advanced display and intelligent interface technologies for C^{3}I. The focus scenario was used to motivate the selection of needed hardware and software component technologies, including wireless pointing systems, spoken language understanding systems, digital video transport, collaborative infrastructures, and audio and video teleconferencing. The key issues in each category were presented, and a significant number of alternatives in each category were reviewed. The resulting selection of laser pointer tracking, Sandy Pentland's vision-based gesture recognition system, BBN's HARK spoken language understanding system, MPEG 1 video support, TASC's CSCW infrastructure (combined with a potential future upgrade to InSoft's Communiqué!), and MBONE tools for audio and video conferencing was motivated by performance, features, functions, availability, and compatibility with existing hardware at the sites involved with the project.

Integration of these basic capabilities is now complete. Substantive upgrades were made to the core communications components of the CSCW architecture. These upgrades were needed to support the novel interface technologies being investigated by VIEP as well as to maintain component abstraction and software reusability. Multi-user tools can now be "plugged into" our CSCW infrastructure and into other infrastructures (such as InSoft's Communiqué!) with greatly reduced effort. A conference scheduling and initiation facility has been incorporated into the VIEP demonstration suite that allows users to create collaborative sessions and notifies conferees of conferences regardless of their location within the network

Several collaborative tools have been incorporated into the conferencing system. These include a high-performance collaborative image viewer, a shared audio player, a collaborative MPEG 1 video player, and the DIVA streaming video player. The image viewer has been enhanced with speech input for image navigation. Video-based gesture recognition software has also been incorporated into the image viewer. Users can control a telepointer on the collaborative viewer using only hand motions. Iconic overlays were also added to the image viewer. The icons are used to represent battlefield elements in a C^{3}I scenario. The icons can be animated using a simple VCR-like control panel enabling VIEP to be used for demonstrations of mission planning, monitoring, and replanning. Additional forms of wireless system interaction were investigated. Laser pointer tracking was incorporated into the system utilizing a detection camera located behind the large projection screen. The laser pointer can be used instead of the mouse to manipulate the VIEP application. An application for wireless text entry using the laser pointer has also been developed. Finally,

an example of how legacy, single-user applications can be incorporated into VIEP was presented.

With the establishment and configuration of the WAN link between Rome Laboratory and TASC's Reading facility, we have been able to demonstrate the collaborative capabilities between the two sites. In fact, four nodes have been used for VIEP, demonstrating VIEP's ability to efficiently handle a multi-user configuration. The focus scenario described above was used as the basis for a demonstration script (shown in Appendix A) that has been used to highlight VIEP's support for mission planning and real-time replanning. It emphasizes the use of audio and video conferencing as well as collaborative applications for timely dissemination of intelligence data and joint decision making. The script has been used for numerous VIEP demonstrations.

We had developed a collaborative infrastructure featuring wireless interaction capabilities by the end of the first year and a half of the VIEP project. Coupled with the prescribed real-time replanning scenario, the CSCW's intrinsic support for non-baton passing interaction and multipoint-to-multipoint communications, we met many of the year two and year three demonstration objectives by the end of that period.With the addition of a videoconferencing capability, tools for mission scenario playback, and support for multi-way collaboration, we met all of the year three demonstration objectives by the end of the second year. Additional capabilities and new technologies (e.g., other wireless user interface devices, improved mission scenario tools) have greatly enhanced the usability of the system.

VIEP has added considerably to the body of knowledge on collaborative computing in general and on the use of collaborative technologies for C^3I in particular. VIEP has shown that 1) a replicated, distributed approach to collaboration is both effective and scalable when combined with reliable multicasting technology, 2) the combination of wireless interface technologies, large, high-resolution displays, and collaborative computing provides for synergies that can be exploited, not only in the C^3I domain, but also in domains such as medicine and science, and 3) access to collaborative tools is critical for the real-time replanning capabilities needed for interdiction of mobile targets. VIEP has pioneered research efforts into the next generation of command and control systems.

APPENDIX A: VIEP DEMONSTRATION SCRIPT

VIEP was designed to demonstrate how the combination of advanced display and intelligent interface technologies with collaborative tools can provide a powerful C^3I capability. This appendix describes a demonstration scenario utilizing VIEP that shows its use for mission planning and real-time replanning.

A.1 Preparations

Network voice (VAT) and video (VIC) communications should be up and running between the demo site and the other collaborators. The voice and video channels allow exchange of verbal and visual cues to the outside collaborators so that they essentially do most of the UI manipulations. Note that only one collaborator (two presenters total) is really necessary to do this demo although the script is written for three presenters. The gesture recognition system should also be calibrated, and the HARK system should be active, with the demonstrator wearing the wireless microphone.

The primary assistant should have a CSCW conference configured and ready to launch, and be waiting for the demonstrator to complete his introduction. All collaborators should have their CSIS Registration Interface tools running and iconified.

A.2 Demonstration Dialog and Presentation Sequence

[Preparation]

- start xgraffiti and LaserPen
- start MpegNetServer on one machine
- start pscene on one machine

[Demonstrator]

- runCSCW
- Join VIEP: Audio and Video
- Join Visual Information Environment Prototype

[Assistant 1]

- runCSCW
- Join VIEP: Audio and Video
- Join Visual Information Environment Prototype

[Assistant 2]

- runCSCW
- Join VIEP: Audio and Video

[Demonstrator] "As a result of the interim peace agreement brought about by NATO's Operation Deliberate Force on 30 August, a temporary stand down from the no-fly resolution has been made to allow bringing aid and relief efforts into Sarajevo. However there is a strong indication that this is an uneasy peace, and intelligence sources indicate that the Serbs are rebuilding their forces despite the cease fire. The demonstration scenario begins with the late breaking intelligence information being brought to our attention at the command center."

[Assistant 2]

- Join Visual Information Environment Prototype
- Drop HARKImageViewer, ImageViewer, ImageViewer, Movie, XAudio, and LiveVideo from ToolBox onto Bulletin Board
- Drop CNN from Jukebox onto Bulletin Board
- Drop uav-hawk, sara, hawk, plane1, and plane2 from VideoStore onto Bulletin Board
- Drop CNN from Channels onto Bulletin Board
- Drop air1, air2, air3, bosnia, and sara from Photo Album onto Bulletin Board

[Demonstrator] "We see here the beginning of a collaborative session that represents a series of interactions with various command and operations components. This first exchange is nominally between intelligence and command centers. The scenario begins with an intercepted radio transmission from Serbian forces."

[Assistant 1]

- Drop CNN onto XAudio
- Open XAudio
- Play XAudio

[Demonstrator] "Assistant one, rewind the audio clip to 23 seconds and play the message"

[Assistant 1]

- Scroll to 23 seconds
- Play XAudio

[Demonstrator] "Obviously this is a recorded clip from CNN rather that a real CommInt intercept; however, since this was captured from a live transmission, there obviously is no reason it couldn't have been a real intercept. The key point here is that digital audio can be stored, transmitted and shared. Let's move forward with the notion that the intercept really conveyed that the Serbs have seized control of the airport, and that they are planning to land a transport at Sarajevo within an hour."

[Demonstrator] "We are able to confirm this information using civilian television broadcasts that intelligence has been monitoring. Here we show a tool for viewing live streaming video from a server located elsewhere in the network."

[Assistant 1]

- Drop CNN onto LiveVideoPlayer
- Open LiveVideoPlayer

[Demonstrator] "These intercepts indicate that the transport will be loaded with munitions destined to resupply the decimated emplacements at Lukavica, Ilijas, and Hadzici."

[Assistant 1]

- Drop air1 onto HARKImageViewer
- Open HarkImageViewer
- Zoom Level 1
- Move to left bottom corner

[Demonstrator] "The transport is a time critical target that needs to be attacked before the munitions can be unloaded and distributed. Overhead imagery provide by National Sources indicates that the transport will likely be stored and unloaded in a hangar near the airfield."

[Assistant 1]

- Move telepointer around the hangar that doesn't get blown up
- Draw circle around hangar

[Assistant 1] "I believe that the transport is in this hangar."

[Assistant 2]

- Choose yellow color
- Move telepointer around the hangar that gets blown up
- Draw square around hangar

[Assistant 2] "The more likely of locations to store the transport is not there but in this hangar."

[Assistant 2] "Demonstrator, here are the possible storage areas."

[Assistant 2]

- Move telepointer around both hangars.

[Assistant 2] "This is the building that will most likely be used to store the transport for un-

loading, and is our primary target."

[Assistant 2]

- Draw arrow sketch pointing to the hangar that gets blown up

[Demonstrator] "I'd like to see this at higher resolution."

[Demonstrator]

- Zoom in
- Zoom in
- Scroll North
- Page West
- Scroll North
- Halt

[Demonstrator] "Assistant one, enhance the contrast and brightness to give me a clearer picture."

[Assistant 1]

- Enhance contrast to 65
- Enhance brightness to 65

[Demonstrator] "You can set the levels back now."

[Assistant 1]

- Reset contrast to 50
- Reset brightness to 50
- Zoom Level 1

[Demonstrator] "So this is the primary target."

[Demonstrator]

- Simulate a button click with LaserPen
- Move LaserPen around yellow square
- Simulate a button click with LaserPen
- Use xgraffiti to type "Primary"

[Assistant 1]

- Type text "Secondary Target"

[Demonstrator] "Let's ready a mission and take out the transport before the munitions get unloaded and distributed."

[Demonstrator] "At this point we begin planning a fast strike out of Aviano to destroy the transport while it is still loaded and on the ground. We begin by examining overhead imagery of Sarajevo with Air Operations out of Aviano.'

[Assistant 1]

- Drop sara onto HARKImageViewer
- Zoom Level 1
- Move to over city

[Demonstrator] "For flight planning purposes, this imagery would be sufficient for mission preparation if the terrain wasn't so rugged. However, since this is Sarajevo, we use the third party software product MUSE to generate a perspective view on the airport. The output is made available to the remote collaborators via an AIMS image server."

[Assistant 1]

- Create a perspective using pscene
- Use xst to place new image icon on Bulletin Board
- Drop new image icon onto second ImageViewer
- Open ImageViewer
- Close ImageViewer

[Demonstrator] "For added emphasis we've created an IPT terrain rendering over DTED to assist in mission rehearsal."

[Assistant 1]

- Drop sara onto Movie
- Open Movie

[Demonstrator] "Here is an actual sequence of IPTs of Sarajevo over DTED. On the fly rendering takes far more computational power that we have here, however we have prerendered this sequence and stored it as an MPEG video."

[Demonstrator] "Assistant one, run this fly through."

[Assistant 1]

- Play Movie

[Demonstrator] "Thank you. Wait. Could you stop this?"

[Assistant 1]

- Stop Movie

[Demonstrator] "Move forward a few frames."

[Assistant 1]

- Frame Forward

[Demonstrator] "Okay. Let it run."

[Assistant 1]

- Play

[Demonstrator] "Because of the terrain, the location of the airport, and the time criticality of the target, the mission must approach the airport from the south. Unfortunately, Intelligence also indicates that the Serbs have been redoubling their air defenses at Mostar since the August 30 raid, and the usual southern entry is extremely risky."

[Demonstrator] "It is decided to launch a coordinated strike against the air defenses at Mostar using a tomahawk cruise missile launched from the Roosevelt carrier group carrying a conventional warhead. The attack is to take place as soon as possible, coordinated with the time that National Assets are over the target area. Here we begin tracking the actual mission."

[Assistant 1]

- Drop bosnia onto HARKImageViewer
- Open ControlTower
- Open bosnia.ct
- Zoom Level 3
- Move bottom right corner
- Hide Lines

[Demonstrator] "Since removing the air defenses at Mostar is a critical risk reduction factor, we have tied a camera to the tail fin of the tomahawk to help evaluate the success of failure of the mission. Let's play to the time of impact of the tomahawk then freeze frame."

[Assistant 1]

- Play Control Tower
- Pause Control Tower after tomahawk disappears
- Drop hawk onto Movie
- Seek to 20
- Play Movie
- Stop Movie
- Play Control Tower
- Pause Control Tower after UAV is past target area

[Demonstrator] "As you can see, the mission appears to be a success, however other analysis folks aren't convinced. Let's review that last couple of seconds. Since time is running out, we need to make a call: continuing along this corridor with the air defenses in place put the mission in jeopardy and the chance of the munitions getting unloaded and distributed increase greatly. Fortunately, an Unmanned Aerial Vehicle was scheduled to fly over the air defense site, and the video has just now become available."

[Assistant 1]

- Drop uav-hawk onto Movie
- Play Movie
- Stop Movie
- Play Control Tower
- Pause Control Tower before SA3 hits plane

[Demonstrator] "Fortunately we got the information in time to redirect the mission. The contingency plan is to fly further south and ingress through an alternative corridor between Mostar and Gorazde. Hopefully the added delay doesn't hamper mission effectiveness. Assistant two, draw the contingency plan."

[Assistant 2]

- Show Lines
- Draw the two redirected F15 flight paths
- Hide Lines

[Demonstrator] "Assistant one, play the mission"

[Assistant 1]

- Play Control Tower
- Pause Control Tower after both F15s are past the target area

[Demonstrator] "As you can see, an SA3 was launched against the mission however the F15s just barely penetrated the air defense perimeter when they were redirected. The SAM loses acquisition and falls harmlessly out to sea."

[Assistant 1]

- Drop air2 onto ImageViewer
- Open ImageViewer

[Demonstrator] "Imagery from National Sources becomes available 15 minutes after mission and reveals that all building except the primary hangar have been successfully destroyed.

However, none of the buildings show the characteristics of a massive munitions explosion. Hence, we haven't succeeded in destroying the transport."

[Demonstrator]

- Simulate a button click with LaserPen
- Move LaserPen around yellow square
- Simulate a button click with LaserPen

[Demonstrator] "Assistant two, lets launch two sorties from the Roosevelt 5 minutes apart against the remaining building."

[Assistant 2]

- Show Lines
- Draw two F18 flight paths 5 minutes apart
- Hide Lines

[Demonstrator] "Assistant one, play the mission to just after the first F18 sortie starts its egress, then freeze frame."

[Assistant 1]

- Play Control Tower
- Stop Control Tower after the first F18 is past the target area
- Drop plane1 onto Movie
- Play Movie
- Stop Movie

[Assistant 1]

- Drop air3 onto ImageViewer

[Demonstrator] "The first F18 sortie reports successfully damaging the remaining building, yet no massive explosion was detected there, either. Gun camera video on egress reveals that the transport has left the hangar and is taxiing towards the runway attempting to leave the area. The information is relayed in real time to the second sortie, which diverts its attack to the departing transport."

[Assistant 1]

- Play Control Tower
- Stop Control Tower after second F18 is past the target area
- Drop plane2 onto Movie
- Play Movie
- Stop Movie

[Demonstrator] "Fortunately, the second sortie completed the objective and the transport is destroyed. Subsequent Intelligence imagery confirms that fact that the primary building has also been destroyed."

[Demonstrator] "Peace will continue at least a little longer."

APPENDIX B: UIST'96 PAPER

A paper describing VIEP was presented at the Association for Computing Machinery's User Interface Software and Technology 1996 (ACM UIST'96) conference. The paper was one of 19 papers accepted out of 75 submissions. The full text of the paper is included in this appendix. It can be cited as:

Rohall, S.L. and Lahtinen, E.P., *The VIEP System: Interacting with Collaborative Multimedia,* Proceedings of UIST'96 (November 6-8, 1996, Seattle, WA), pp. 59-66.

The VIEP System: Interacting with Collaborative Multimedia

Steven L. Rohall *Eric P. Lahtinen*
TASC
55 Walkers Brook Drive
Reading, MA 01867-3297 USA
+1 617 942 2000
{slrohall, eplahtinen}@tasc.com

ABSTRACT
This paper presents a survey of the Visual Information Environment Prototype (VIEP), a system which demonstrates the next generation of Command, Control, Communication, and Intelligence (C3I) systems. In particular, VIEP provides a novel integration of user interaction techniques including wireless input and large-screen output to facilitate the task of collaborating with media such as large images, audio, and video. The prototype has been implemented and demonstrated over both local and wide area networks.

KEYWORDS
CSCW, Collaboration, Multimedia, Wireless Interfaces

INTRODUCTION
This paper presents a survey of the Visual Information Environment Prototype (VIEP). VIEP has been designed and implemented to demonstrate how advanced user interaction techniques coupled with high-performance collaboration can improve decision making. In particular, VIEP is targeting the Command, Control, Communication, and Intelligence (C3I) environment.

The Problem
The C3I environment is a demanding one [3]. Traditionally, the heart of C3I has been the "situation room" which consists of a room of operators receiving intelligence reports from field personnel. Information is collected manually on large, static displays using grease pencils. Commanders, distancing themselves somewhat from the data displays, try to decipher the big picture and then issue orders accordingly.

However, the nature of conflict is changing. Future military actions are likely to be limited-scale and theater-oriented, necessitating the move to distributed decision making where field personnel communicate directly with the commanders as well as each other. At the same time, the amount and types of intelligence data have increased. Surveillance photos, weather data, audio intercepts, video from unmanned air vehicles, and more must be acquired and analyzed to make the appropriate decisions. Commanders are more likely to be overwhelmed rather than aided by this additional information.

Current computer systems for C3I have only automated the traditional ways of doing things [12]. For example, large screen computer displays have replaced the maps painted onto glass plates and computer networks are used extensively for data collection. However, the computer technology has not been used to 1) aid in the decision making process, especially when that process is distributed, 2) facilitate the analysis of large amounts of multimedia data, and 3) make the computer systems more accessible to all users, especially high-level commanders who have been reluctant to use keyboards, mice, and other input devices in the past.

The Approach
To address these problems, VIEP is using emerging technologies in user interaction, display, and collaborative computing to create a new type of C3I system. Situation rooms, connected via high-speed networking, will be dominated by large, rear-screen projection displays. Commanders and operations personnel will manipulate data using a variety of wired and wireless technologies. Field personnel are brought into the decision-making process

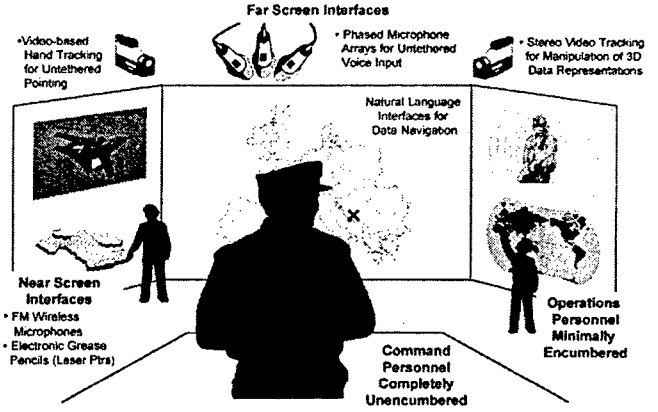


Figure 1: The VIEP Vision

This work was sponsored in part by the USAF (Rome Laboratory), contract number F30602-94-C-0072.

through teleconferencing and collaborative applications. This vision is depicted graphically in Figure 1.

There has been progress in the necessary component technologies. User tracking systems and voice recognition have been made faster and more accurate. Large screen displays have been made larger and less expensive. Increases in network bandwidth and transmission protocols have enabled more rapid transmission of imagery. Collaborative computing has reached a point where it can be used to support network-based conferencing and interaction.

While much has been accomplished on the component technologies, very little has been done to examine their combination and interaction. VIEP brings together these pieces to form an innovative environment that will greatly facilitate the control and understanding of data.

TECHNOLOGY COMPONENTS

The philosophy throughout the project has been to develop technology only where needed, to integrate existing tools and technology whenever possible, and to implement the system within a fairly tight budget. In the survey of VIEP technology that follows, these choices will become clear. There are three main component technology areas: collaborative computing, wireless user input, and large display output.

Collaborative Computing

The Computer-Supported Cooperative Work (CSCW) component provides the basic infrastructure in VIEP for integrating the various interface technologies and providing the collaborative C3I capabilities. The system has to support a large number of widely-distributed users in an efficient manner. It also has to provide good performance in spite of potentially-large network latencies and the use of large sets of multimedia data. Finally, the system has to be extensible and provide an API so that new tools can be added without needing to modify large parts of the system.

Although there are several commercial offerings of desktop conferencing systems, none met our design requirements of scalability to a large number of users, good performance when working with large datasets (100 Mbyte images, for example) over wide area networks, flexibility in supporting a wide range of user interactions including side conferences

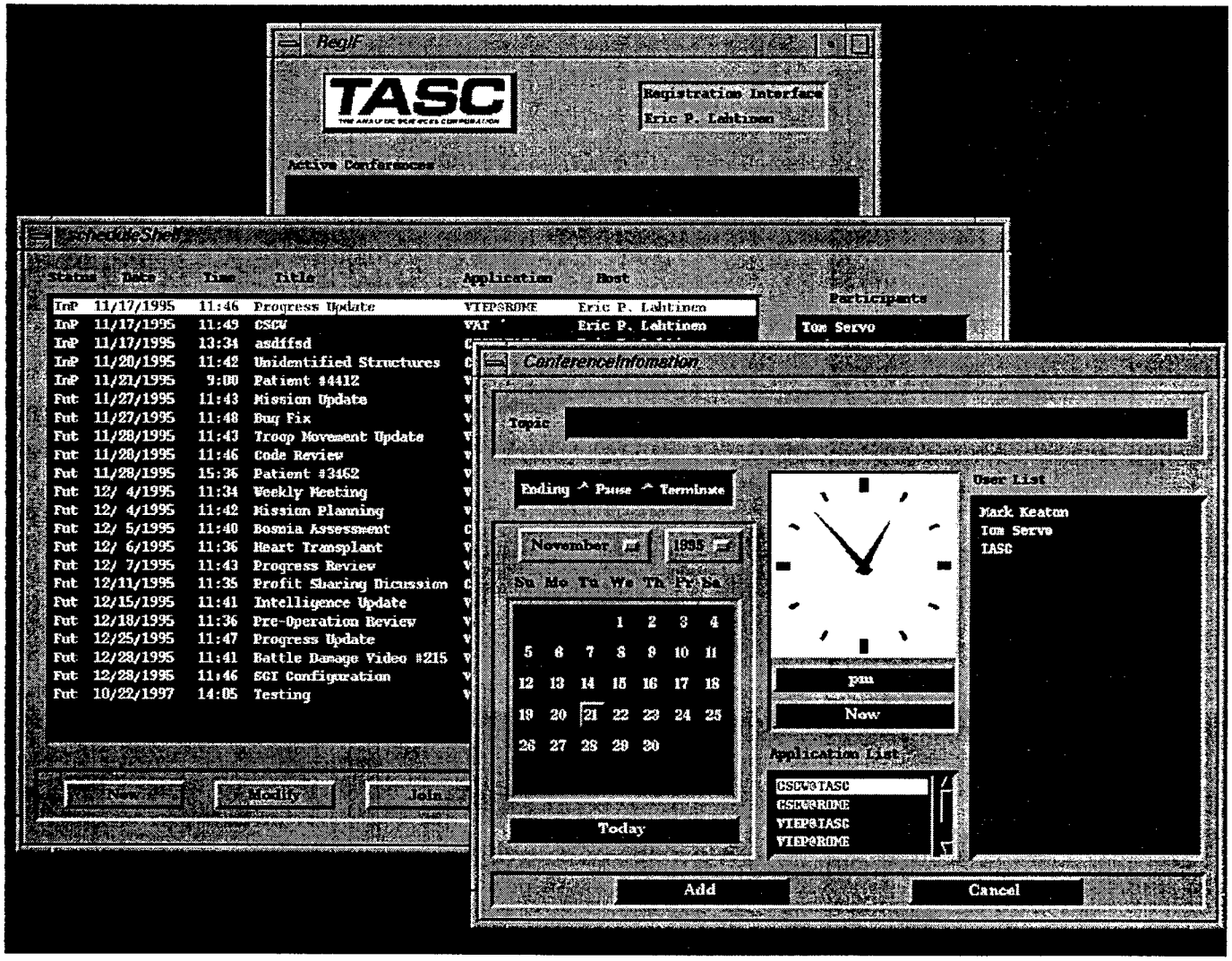


Figure 2: Conference Scheduling and Initiation

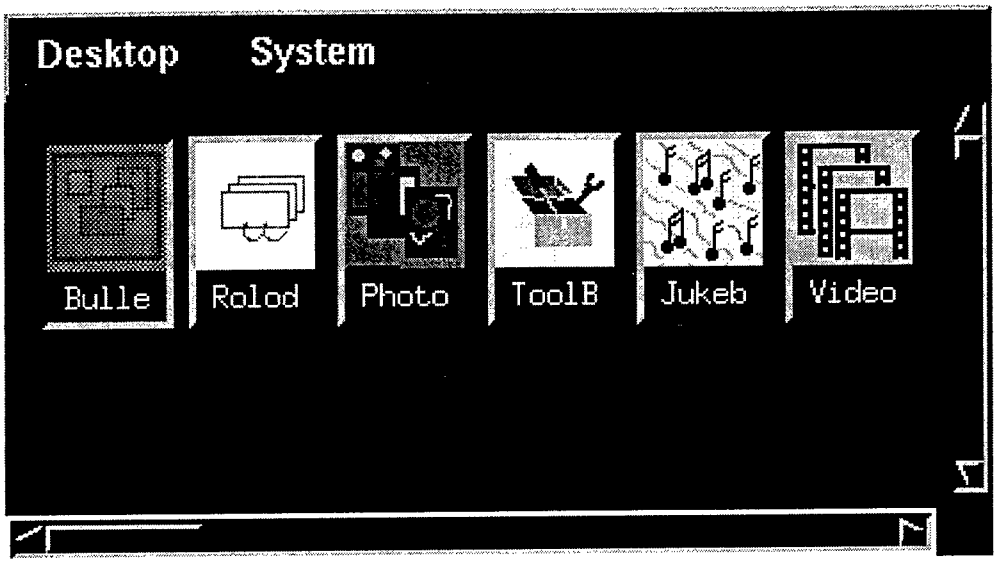


Figure 3: MultiTalk Desktop

and late comers, and extensibility. We decided to use TASC-developed software instead. There are two components to our collaborative software environment: a Conference Scheduling and Initiation System, CSIS, and a multicast-based collaborative infrastructure, MultiTalk.

Conference Scheduling and Initiation. TASC's Conference Scheduling and Initiation System (CSIS) supports the creation of collaborative conferences [7]. It allows users to schedule conferences, specify which application tools will be used during the conference, and launch the applications at the appropriate time.

The interface to CSIS is shown in Figure 2. The top window in Figure 2 is the Registration Interface which shows the user name as well as any active conferences. From this window, the user may join a conference to which he has been invited or schedule a new conference. When the user schedules a new conference, the second window on the left of Figure 2 is used. This window shows the entire set of conferences managed by CSIS along with an indication of whether they are in progress or scheduled for future execution, the date and time of the conference, the conference title, which application tools are going to be used in the conference, and who created the conference. To create a new conference, the window shown in the lower right of Figure 2 is used. The conference creator specifies the date and time for the conference, the application, and the users to participate in the conference. Conferees are specified by name; CSIS takes care of finding participants regardless of their network address.

For VIEP, the two main applications that the user may specify when creating a conference are Teleconferencing and TASC's MultiTalk. The Teleconferencing application allows conferees to see and hear each other. MultiTalk is a TASC-developed conferencing substrate used for implementing multi-user applications.

Teleconferencing. To provide audio and video teleconferencing, we have incorporated the MBONE tools, *vic* (for video) and *vat* (for audio), into VIEP [8, 2]. These tools provide adequate audio and video conferencing and have the added advantage that they utilize multicast. As a result, they are very network efficient and allow the system to be scaled to a large number of users.

TASC MultiTalk. TASC MultiTalk implements a desktop metaphor for the sharing of data among conference participants [15]. The main desktop window is shown in Figure 3. It contains easily-accessible collections of data items including a photo album of imagery, a jukebox of audio clips, and a video library of video clips. These are the data items that the collaborators will be manipulating. There is also a toolbox on the desktop which contains collaborative tools for manipulating the data: a shared image viewer, a collaborative audio player, and a shared MPEG-1 video player. The final item on the desktop, and perhaps most important, is the bulletin board.

The bulletin board represents the shared state of the conference. A typical bulletin board from a conference in session is shown in Figure 4. Everything on the bulletin board is seen and can be manipulated by everyone in the conference. Included on the example bulletin board are icons of the conference participants, several data items (an image "1airp," an audio clip "UNPol," and a video clip "Hit_P"), and three collaborative viewing tools (an image viewer "TalkI," an audio player "Audio," and a movie player "Movie").

Associations are made between conferees, viewers, and data items by dragging and dropping the respective icons. In Figure 4, an association has been made between the audio player, the "UNPol" audio clip, and the two conference participants. When either participant double clicks the audio player, both will hear the clip. Associations allow a broad range of collaborative activities. They can be used to support subconferences or side discussions, for example.

Collaborative Tools. As described above, there are three collaborative tools available in MultiTalk: a high-performance image viewer, an audio player, and a video player. The audio player utilizes built-in audio capabilities of

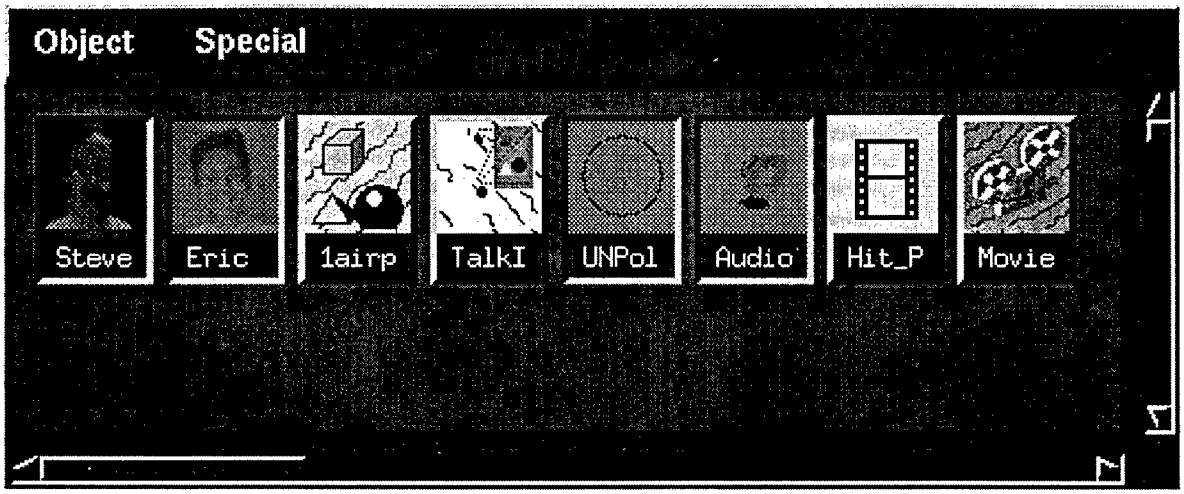


Figure 4: MultiTalk Bulletin Board

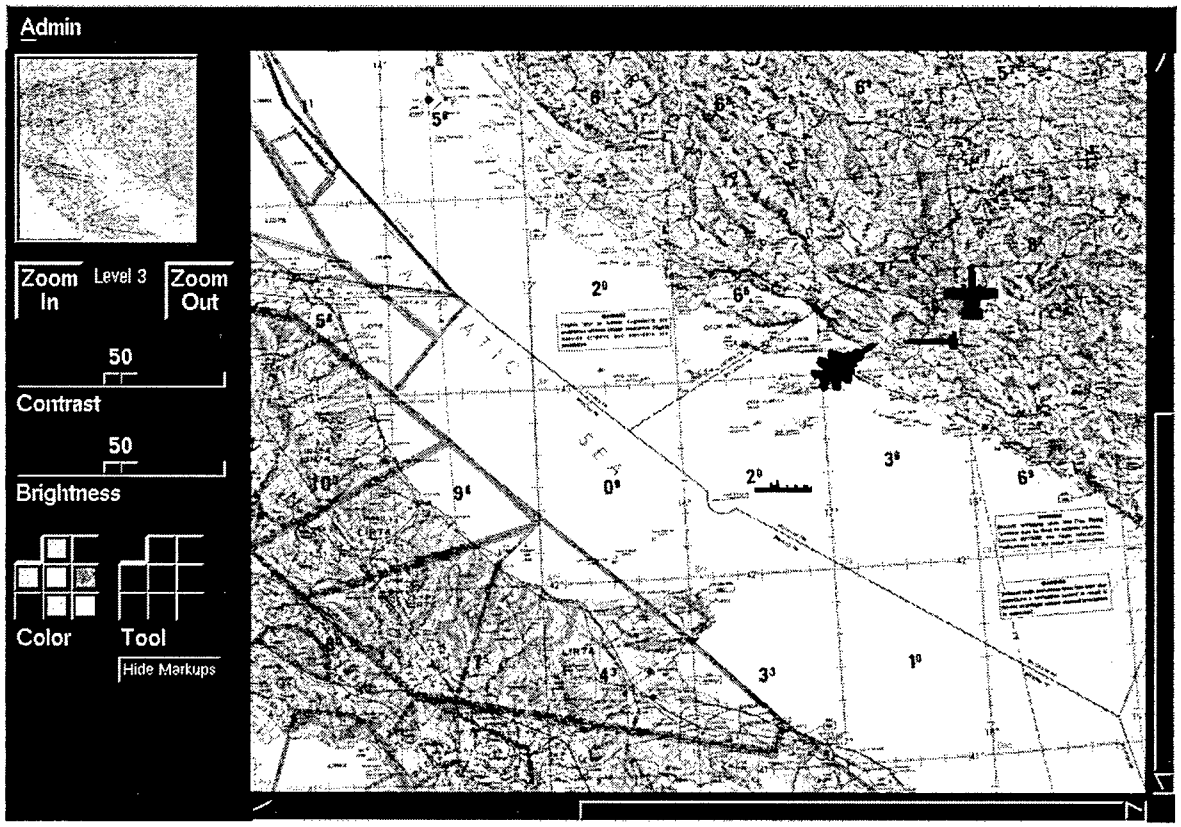


Figure 5: Collaborative Image Viewer

the workstations used for the prototype. The video player is based on the MPEG-1 player developed at Berkeley [9]. In both cases, state changing commands such as "play" and "stop" are sent when a user starts or stops an audio or video playback. Replicated copies of the collaborative tools on the other conferees' machines then execute the commands. We have been able to achieve synchronization among the users despite network latencies even with replicated applications. This is achieved through specialized synchronization primitives described more fully in [15]. In the present system, the audio and video players operate on local copies of the data items. This was done purely to save network bandwidth between the collaborators. We are currently working on the modifications necessary to handle audio and video streamed from one machine in the network.

The collaborative image viewer, shown in Figure 5, was originally developed as part of the IMACTS project at TASC [13]. The viewer is able to handle images of extremely large size when they are stored in the AIMS image format (a tiled, hierarchical image format based upon TIFF). AIMS images are organized as multi-resolution pyramids. The viewer shows the user a low-resolution overview of the image and

allows the user to zoom in to particular areas of interest. At any particular resolution, only the tiles needed to fill the high-resolution portion of the viewer are fetched. Additional tiles are fetched as needed when the user scrolls and zooms the image. At its highest resolution, the image shown in Figure 5 is only about 6000 pixels on a side; however, the viewer has handled images well in excess of 1 Gbyte of data.

Under VIEP, the image viewer was made collaborative. Image loading, panning, scrolling, and zooming functions are shared among all of the viewers in a conference. In addition, the image viewer has the capability to fetch data from a server in the network or use a local copy. The viewer was also augmented with telepointers for collaborative use. Any number of telepointers can be shown and manipulated simultaneously. The user can select the color for his pointer, and when he presses a mouse button in the high-resolution portion of the viewer, the pointer location is shown in all of the viewers in the specified color.

Finally, the image viewer has been enhanced with animated, iconic overlays. In Figure 5, these are used to represent planes and ships. The icons can be moved under programmatic control, allowing the viewer to be used for situation monitoring or "what if" planning.

Multicast. Like the MBONE tools, MultiTalk makes extensive use of multicast in order to use the network most efficiently. The underlying communication among MultiTalk desktops is via RAMP, the TASC-developed Reliable Adaptive Multicast Protocol [14]. RAMP combines the network efficiency of multicast with the programming convenience of reliability.

Wireless Input Devices

The second key technology component in VIEP is the use of wireless input devices. Commanders have been reluctant to use technology that requires them to be wired to a computer. Keyboards and mice, let alone data gloves and virtual reality helmets, are deemed too odious. In addition, during time of crisis, commanders should not have to search for a mouse to access the necessary information. In fact, there may be more people in the main VIEP situation room than there are keyboards and mice.

One of the challenges for VIEP has been to find alternative user interface technologies that 1) are more natural to use, especially for high-level commanders, 2) allow for a broad range of interactions with data, including 3-dimensional data, and 3) support multiple users in front of the large-screen display. The current VIEP system uses two technologies to address these problems: video-based gesture recognition and speech recognition.

Gesture Recognition. People naturally want to use their hands when interacting with systems, both for specifying (selecting) objects and for manipulating those objects. Mice, trackballs, touch pads and screens, and graphics tablets are limited to 2-D and 2.5-D datasets, have generally been wired to the computer, and require a stable base or two hands to operate. Three-dimensional input devices, such as data gloves, space balls, and flying mice, tend to be expensive and bulky when adapted for untethered use.

Increases in computational power and improvements in processing algorithms have made video-based person tracking systems possible. By means of a video camera, the computer tracks the positions of the user's hands. The user needs no special equipment. Chroma keying can improve the accuracy and performance of such systems, requiring the user to wear simple, colored gloves.

We have incorporated a video-based gesture recognition system developed at the MIT Media Laboratory into VIEP [10, 11]. The software detects motion by subtracting a pre-acquired background image from the incoming video stream. The difference image is fed to an image understanding module that infers the location of various body components, allowing hand positions to be isolated.

Initially, the gesture recognition software has been used to control a telepointer on the collaborative image viewer. Figure 6 shows the integrated system in operation. On the left side of Figure 6, the user is standing in front of a workstation with a camera on top of its monitor. The window in the middle of Figure 6 shows the output of the image understanding module; this is what the system "sees" after the pre-acquired background image is subtracted from the live video stream. Ultimately, the hand position is used to control a telepointer on the collaborative image viewer. In Figure 6, the telepointer is the small square under the "E" in the label "CRES" in the upper left of the image.

Stereo person tracking, utilizing two input cameras, is also possible. We plan to incorporate this improved software into VIEP in the future.

One shortcoming of video-based gesture recognition is that it is only suitable for standoff operation. As the user approaches the camera (which will in operation be mounted above a large, rear-screen projection system), he either obscures the background, thereby confusing the background subtraction module, or he leaves the camera's field of view totally. Stereo tracking only exacerbates this problem, since the user must be visible to both cameras. This is suitable for a commander, standing back and surveying the situation, but it is not appropriate for the other operators who need to work more closely to the screen.

We have determined that a different type of wireless pointing input is needed for near-screen operation. However, commercially-available technologies for wireless pointing do not satisfy our requirement of supporting multiple, near-screen users. We are looking into using laser presentation pointers coupled with laser detection technology behind the rear-screen projector as an analog for the grease pencils currently used by operations personnel. The electronic grease pencil would be similar in operation to the pen used in the Liveboard system [4].

Speech Recognition. At its best, video-based recognition is still fairly low resolution. Although it is possible to track the hands, it is difficult, for example, to localize the fingers. This makes it difficult to use any sort of sign language for selecting and manipulating data items. We have decided to augment the wireless gesture recognition with speech recognition to overcome this problem. Speech recognition

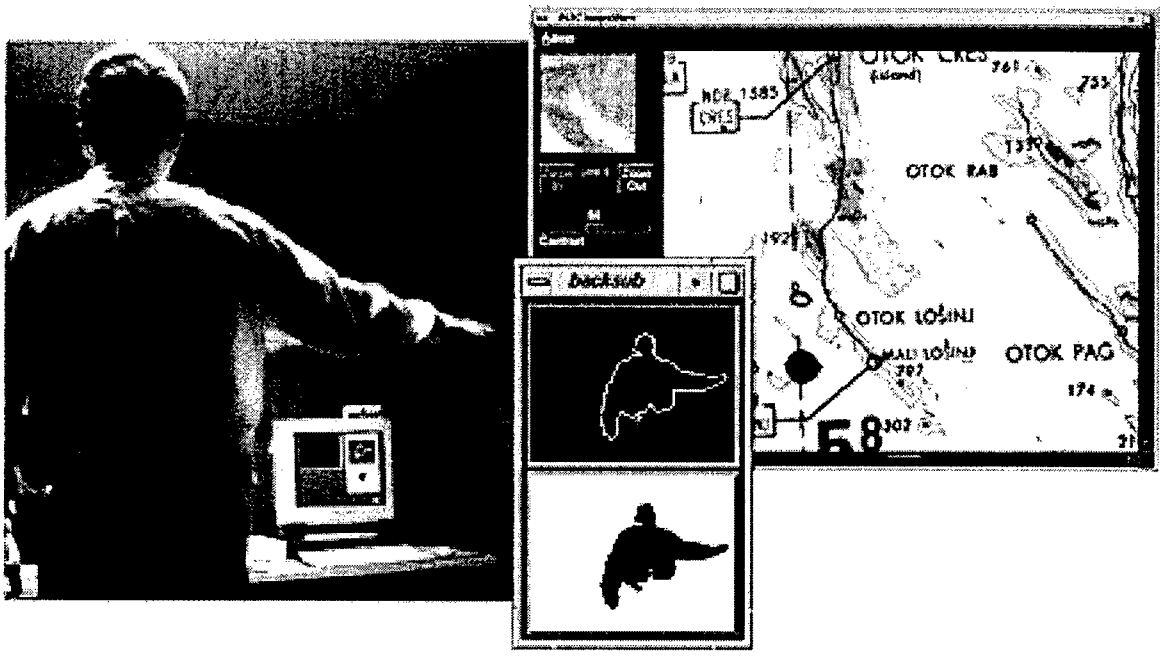


Figure 6: Gesture Control of the VIEP System

naturally complements gesture recognition, allowing people to point and talk to get their work done [1, 6].

The performance of speech recognition systems has been greatly enhanced through the inclusion of natural language processing. By relying on phrase recognition rather than individual word recognition, accuracy of these systems has increased greatly. The larger benefit of spoken language understanding systems, however, is gleaned through the resulting interpretation of the sentence or phrase, in that the meaningless or extraneous words that naturally occur in speech are automatically removed, and physically different phrasings of essentially the same request can be mapped to the same set of actions.

Although our initial requirements were only for limited word recognition to implement select functions, our eventual goal is to incorporate more functional spoken language understanding into VIEP. There are many commercial and research speech recognition systems available. We chose to integrate BBN's HARK™ Recognizer speech recognition software into VIEP [5]. HARK performs continuous speech recognition and can be configured for speaker independent recognition.

At this time, speech recognition is used to control the panning, scrolling, and zooming functions of the collaborative image viewer. By speaking commands such as "scroll north," "go south," or "zoom in" the user is able to control image position and magnification. By repeating a direction, the user increases the scrolling speed in that particular direction. Both spoken commands and mouse input can be used interchangeably. In addition, incoming commands from remote collaborators are also processed at the same time.

Wireless and Traditional Interfaces. We do not expect that we will totally eliminate the use of conventional input devices in VIEP. Indeed, we have designed the system so that it works equally-well with either wired or wireless input (or both). We anticipate that trained operators, such as those that work in the situation rooms today, will continue to be the primary people interacting with the system. That class of user will certainly use traditional input devices.

The MultiTalk subsystem supports the notion of varied capabilities through a "plug in" architecture that allows users to pick which capabilities they would like to use. In the main situation room, speech and gesture recognition may be plugged in for user input. At a remote site, there may not be enough room for gesturing or the environment may be too noisy for speech recognition. These users do not need to worry about those components and can interact with the system through traditional interfaces.

Large Screen Output

The third key technology component in VIEP is the use of advanced display devices. Although the amount of data available to users has increased dramatically, this increase has not been followed by a corresponding increase in the number of and effectiveness of tools to aid users in the understanding the data. A simple method for helping users manage the increased data volume is to provide larger display areas, both in terms of resolution and in terms of sheer physical size.

In essence, there are two types of collaboration within the VIEP system. The first, described above, is the collaboration with remote personnel using networks and teleconferencing. The second is the collaboration among the commanders and operators co-located in one of the situation rooms. These users need to be able to see and manipulate each other's data. One user may be working in isolation with some data, but eventually, the analysis of that data will need to be presented

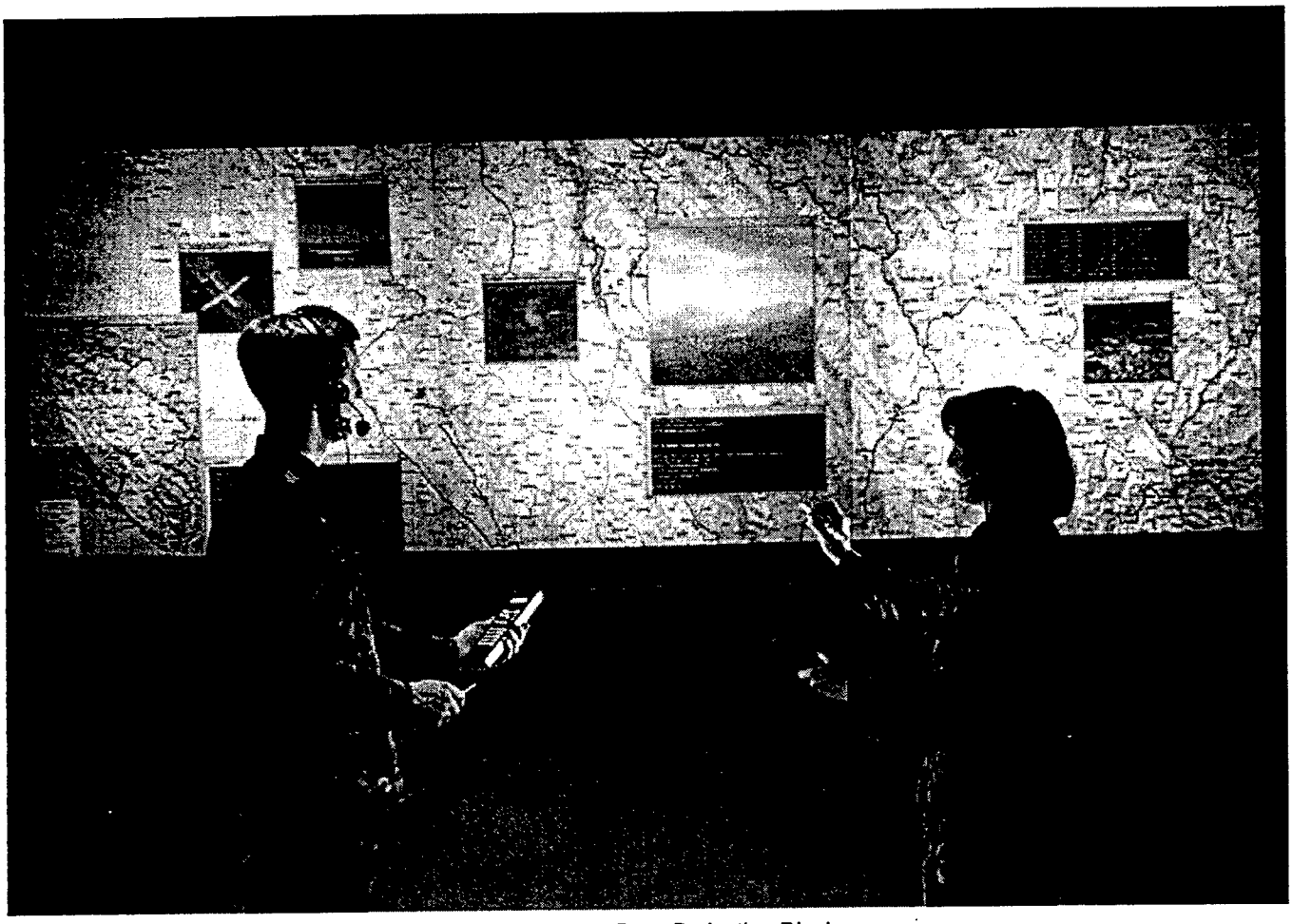


Figure 7: Large, Rear-Projection Display

to the group. A large screen display provides enough space to allow individuals to work close to the screen on a set of data. At the same time, it affords the legibility for a group of users to stand back and discuss that data.

The VIEP sponsor has constructed such a display for the use of this and other projects. A photograph of this display is shown in Figure 7. The display consists of three, horizontally-tiled video projectors each driven at 1200 by 1600 pixels to produce a tiled display with a total resolution of 1200 by 4800 pixels across a screen area of 40 inches by 120 inches. The high-resolution data projectors are mounted behind a glass screen; rear projection is employed so that users can stand directly in front of the display. The software treats the three monitors as one, large X Window System root window. Each projector actually has a video bandwidth approaching 2000 by 2500 pixels yielding a future display capability of approximately 15 million pixels.

As mentioned previously, the VIEP program is interested in investigating three-dimensional data. There are two standard techniques for producing 3D in a projection system. The first is to use two projectors polarized orthogonally to each other. The users wear correspondingly polarized glasses to receive a different image on each eye. It is critical in this configuration that the projection screen be non-depolarizing. The second way is to alternately project left and right images. The users wear active glasses that block the eyes alternating in synchrony with the projector. If the switching speed is fast enough, "cross talk" between the two channels is minimized. Both options are quite costly; VIEP is likely to use more limited 3D display capabilities in the near future.

STATUS AND FUTURE WORK

VIEP is a 3.5 year program. We have just completed our second year of the development and integration effort. The system discussed in this paper is demonstrable and has been used over both local and wide area networks. Feedback from the program sponsor has been very positive.

There are several areas where we are actively developing the system. First, we are investigating additional types of wireless input devices, particularly for near-screen use. As described previously, we are developing an electronic "grease pencil" using laser presentation pointers and video cameras located behind our projection screen. In addition, electric field sensing technology looks promising as a means of interacting with the system [16].

Second, we expect to improve upon the wireless interface technologies already present in VIEP by incorporating stereo person tracking as well as increasing the amount of the system that is controlled via gesture recognition. We would also like to increase the type of system operations that can be

performed via speech input as well as expand the system's lexicon to include more complex phrases in addition to simple commands.

We also expect to further investigate the use of audio output for VIEP. At present, audio output is only used for the playback of audio clips. Audio could prove valuable in providing system feedback, especially in the case where the users are standing back from the display and more traditional textual feedback may not be readily visible.

Work is also underway to incorporate audio and video streaming technology into the system. This is particularly important for information sources that acquire data in real-time. There may not be adequate time to capture the data item and then ship it between the potential collaborators.

Finally, as mentioned previously, we are interested in adding three-dimensional data and viewing to VIEP.

CONCLUSION

The VIEP system is a synergistic combination of what have been, up to now, separate technologies: wireless interfaces including gesture and speech recognition; extremely large, high-resolution displays; and collaborative computing encompassing traditional audio and video teleconferencing as well as specialized multi-user applications. While developed to support command and control operations, we feel that such a system is applicable in other environments as well. In particular, any environment where there is a lot of multimedia data that needs to be communicated among distributed users, such as medicine or science, is a candidate for a VIEP-like system.

ACKNOWLEDGMENTS

Mark Keaton implemented large portions of CSIS. Steve Zabele has been instrumental in formulating the design of the VIEP system.

REFERENCES

1. Bolt, R. *Put That There: Voice and Gesture at the Graphics Interface*, Computer Graphics, July 1980, 14(3), pp. 262-270.
2. Casner, S. and Deering, S., *First IETF Internet Audiocast*, ACM SIGCOMM Computer Communications Review, July 1992, 22(3), pp. 92-97.
3. Coakley, T.P., ed. *C3I: Issues of Command and Control*, National Defense University, Washington D.C., 1990.
4. Elrod, S., *et al.*, *Liveboard: A Large Interactive Display Supporting Group Meetings, Presentations, and Remote Collaboration*, Proceedings of CHI'92 (May 3-7, 1992), pp. 599-607.
5. *HARK™ Recognizer Programmer's Guide*, BBN, Cambridge, MA, June 1994.
6. Hauptmann, A.G., *Speech and Gestures for Graphic Image Manipulation*, Proceedings of CHI'89 (April 30-May 4, 1989), pp. 241-245.
7. Keaton, M. and Zabele, G.S., *A Conference Scheduling and Initiation Service for Collaborative Applications*, Proceedings of the Technical Conference on Telecommunications R&D in Massachusetts (March 12, 1996, Lowell, MA), pp. 351-361.
8. McCanne, S. and Jacobson, V., *vic: A Flexible Framework for Packet Video*, Proceedings of MultiMedia'95 (November 5-9, 1995, San Francisco, CA), pp. 511-522.
9. Patel, K., Smith, B.C., and Rowe, L.A., *Performance of a Software MPEG Video Decoder*, Proceedings of MultiMedia'93 (August 1-6, 1995, Anaheim, CA), pp. 75-82.
10. Pentland, A., *Smart Rooms*, Scientific American, April 1996, 274(4), pp. 54-62.
11. Pentland, A., *Machine Understanding of Human Action*, Proceedings of 7th International Forum on Frontier of Telecommunication Technology, (November 1995, Tokyo, Japan).
12. Rackham, Peter, ed. *Jane's C4I Systems*, Jane's Information Group, Inc., Alexandria, VA, 1995.
13. Stephenson, T. and Voorhees, H., *IMACTS: An Interactive, Multiterabyte Image Archive*, Proceedings of the Fourteenth IEEE Symposium on Mass Storage Systems (September 11-14, 1995, Monterey, CA), pp. 146-160.
14. Zabele, G.S., DeCleene, B., and Koifman, A., *Reliable Multicast for Internet Applications*, Proceedings of the Technical Conference on Telecommunications R&D in Massachusetts (March 12, 1996, Lowell, MA), pp. 476-485.
15. Zabele, G.S., Rohall, S.L., and Vinciguerra, R.L., *High Performance Infrastructure for Visually-Intensive CSCW Applications*, Proceedings of CSCW'94 (October 22-26, 1994, Chapel Hill, NC), pp. 395-403.
16. Zimmerman, T.G., *et al.*, *Applying Electric Field Sensing to Human-Computer Interaction*, Proceedings of CHI'95 (May 7-11, 1995, Denver, CO), pp. 280-287.

*U.S. GOVERNMENT PRINTING OFFICE: 1998-610-130-61213